МІЖНАРОДНИЙ ГУМАНІТАРНИЙ УНІВЕРСИТЕТ
Факультет кібербезпеки, програмної інженерії та комп'ютерних наук
Кафедра комп'ютерної інженерії та інноваційних технологій

Пояснювальна записка

до кваліфікаційної роботи
другого (магістерського) рівня

на тему ДОСЛІДЖЕННЯ АНОМАЛІЙ МЕРЕЖНОГО ТРАФІКА DDoS-АТАК

Виконав: студент 2 курсу, групи КТК-2.1
спеціальності 125 Кібербезпека

Севрюков О.В.

Керівник Соловська І.М.

Рецензент Григор'єва Т.І.

Одеса – 2023

ДОВІДКА

кафедри КІ та ІТ про виконану магістерську роботу

студента 2 курсу ФКПІ та КН групи КТК-2.1

Севрюкова Олександра Володимировича

на тему Дослідження аномалій мережного трафіка DDoS-атак

Висновок нормоконтролера кваліфікаційної роботи виконано з порушеннями вимог ДСТУ. Оформлено згідно вимог внутрішнього положення НТУ

Нормоконтролер кан. доцент
(науковий ступінь, вчене звання, посада) (підпис, дата) В.В. Педаш
(і.б. прізвище)

Висновок відповідального за наявність плагіату згідно з сертифікатом ID 1015700948 унікальність роботи підтверджено

Відповідальна особа кан.к, доцент
(науковий ступінь, вчене звання, посада) (підпис, дата) В.В. Педаш
(і.б. прізвище)

Попередня експертиза (захист) магістерської роботи
(бакалаврської роботи чи магістерської роботи)

студ. Севрюкова О.В. проведена "15" грудня 2023 р.
(прізвище і.б.)

Висновки Виконання МР відповідає завданню усі пункти виконано якісно та згідно вимог до оформлення.
Оригінальність роботи: запропоновано дослідження аномалій мережного трафіка DDOS-атак на базі статистичних методів та алгоритмів класифікації машинного навчання для забезпечення необх. точності визначення DDOS-атак.
МР відповідає вимогам до ВКР за заявленою спеціальністю 125 кібербезпека та може бути рекомендована до захисту в ДЕК.

Члени комісії (підпис) к.т.н., доц. Гроза А.І.
(науковий ступінь, вчене звання, посада, прізвище і.б.)

(підпис) кан. доцент Педаш В.В.
(науковий ступінь, вчене звання, посада, прізвище і.б.)

(підпис) вик. каф. КІ та ІТ Цивець О.В.
(науковий ступінь, вчене звання, посада, прізвище і.б.)

МІЖНАРОДНИЙ ГУМАНІТАРНИЙ УНІВЕРСИТЕТ

Факультет кібербезпеки, програмної інженерії та комп'ютерних наук
Кафедра комп'ютерної інженерії та інноваціних технологій
Освітній ступінь магістр
Галузь знань 12 Інформаційні технології
Спеціальність 125 Кібербезпека

ЗАТВЕРДЖУЮ
Завідувач кафедри КІ та ІТ
к.т.н., доц.
_____ Л.Г. Йона
"25" 09 2023 року

З А В Д А Н Н Я
НА МАГІСТЕРСЬКУ РОБОТУ

Севрюкову Олександру Володимировичу

1. Тема роботи: Дослідження аномалій мережного трафіка DDoS-атак
керівник роботи Соловська Ірина Миколаївна к.т.н., доцент кафедри комп'ютерних наук
затверджені наказом закладу вищої освіти від 25 вересня 2023 р. № 1951

2. Строк подання студентом роботи 11.12.2023 р.

3. Вихідні дані до роботи: 1) Класифікація та характеристики DDoS-атак згідно ITU-T рекомендацій ITU-T X.1042 та ITU-T X.1752. 2) Характеристики мережного трафіку та методи детектування DDoS-атак. 3) Алгоритми та методи машинного навчання щодо детектування аномалій мережного трафіку DDoS-атак. 4) Приклади тестових даних трафіку DDoS-атак.

4. Зміст розрахунково-пояснювальної записки
Розділ 1: Огляд технологій та методів виявлення та попередження аномалій мережного трафіку DDoS-атак
Розділ 2: Детектування аномалій мережного трафіку DDoS-атак за допомогою статистичних методів
Розділ 3: Використання алгоритмів машинного навчання щодо детектування аномалій мережного трафіку DDoS-атак

5. Перелік графічного матеріалу (з зазначенням обов'язкових креслень)
Слайд 1 –Класифікація технологій та методів виявлення та попередження аномалій мережного трафіку DDoS-атак

Слайд 2 –Формування вибірок мережного трафіку та характеристики трафіку DDoS-атак
Слайд 3 –Визначення та прогнозування трафіку DDoS-атак
Слайд 4 –Використання алгоритмів класифікації для детектування аномалій мережного трафіку DDoS-атак
Слайд 5 –Аналіз отриманих результатів

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Завдання видав	Завдання прийняв

7. Дата видачі завдання 25.09.2023 р.

КАЛЕНДАРНИЙ ПЛАН

№ з.п	Назва етапів магістрської роботи	Строк виконання етапів роботи	Примітка
1	Вступ	25.09.2023-28.09.2023	Вик
2	Огляд технологій та методів виявлення та попередження аномалій мережного трафіку DDoS-атак	29.09.2023-23.10.2023	Вик
3	Детектування аномалій мережного трафіку DDoS-атак за допомогою статистичних методів	24.10.2023-10.11.2023	Вик
4	Використання методів машинного навчання щодо детектування аномалій мережного трафіку DDoS-атак	11.11.2023-23.11.2024	Вик
5	Висновки та рекомендації	24.11.2023-30.11.2023	Вик
6	Перелік посилань	1.12.2023-5.12.2023	Вик
7	Додаток А. Перелік демонстраційного матеріалів	6.12.2023-8.12.2023	Вик

Студент _____ О.В. Севрюков
(підпис)

Керівник роботи _____ І.М. Соловська
(підпис)

ВІДГУК КЕРІВНИКА

на кваліфікаційну роботу здобувача другого (магістерського) рівня
Севрюкова Олександра Володимировича
на тему: «Дослідження аномалій мережного трафіка DDoS-атак»

Сучасний технологічний розвиток інфокомунікаційних систем та мереж супроводжується необхідністю забезпечення концепції безпеки мережевої інфраструктури від різних атак, в тому числі атак, що призводять до відмови в обслуговуванні DDoS-атак (SYN-Flood, ICMP-Flood, UDP-Flood). Виявлення та попередження аномалій мережного трафіка є актуальною темою дослідження.

В магістерській роботі проведено дослідження аномалій мережного трафіка DDoS-атак на базі статистичних методів та алгоритмів класифікації машинного навчання, таких як, алгоритм логістичної регресії, алгоритм k-найближчих сусідів та алгоритм випадкового лісу.

Здобувач Севрюков О.В. повністю виконав завдання до кваліфікаційної роботи. В процесі роботи здобувач Севрюков О.В. працював самостійно. Графік консультацій не порушувався. Поставлене завдання виконано у повному обсязі. Пояснювальна записка та демонстраційна презентація виконана охайно із дотриманням усіх необхідних вимог.

Під час виконання кваліфікаційної роботи здобувач Севрюков О.В. розібрався з усіма поставленими питаннями та показав уміння користуватись технічною літературою, ставити та розв'язувати дослідницькі задачі.

Кваліфікаційна робота відповідає вимогам до кваліфікаційних робіт другого (магістерського) рівня та заслуговує оцінки «відмінно».

Здобувач Севрюков О.В. заслуговує присвоєння кваліфікації магістр з кібербезпеки за заявленою спеціальністю 125 «Кібербезпека».

Керівник,
завідувач кафедри
комп'ютерних наук
к.т.н., доцент

І. М. Соловська

РЕЦЕНЗІЯ

на кваліфікаційну роботу здобувача другого (магістерського) рівня
Севрюкова Олександра Володимировича
на тему: «Дослідження аномалій мережного трафіка DDoS-атак»

Кваліфікаційна робота здобувача Севрюкова О.В. присвячена питанням дослідження аномалій мережного трафіка DDoS-атак.

У магістерській роботі проведено дослідження аномалій мережного трафіка DDoS-атак. Проаналізовано основні види DDoS-атак та основні характеристики трафіку атак. Запропоновано дослідження аномалій мережного трафіка DDoS-атак на базі статистичних методів та методів машинного навчання, таких як, метод логістичної регресії, метод k-найближчих сусідів та метод випадкового лісу. Результати дослідження дозволять забезпечити необхідну точність визначення DDoS-атак.

Здобувач Севрюкова О.В. має достатню теоретичну підготовку та добре володіє матеріалом. Кваліфікаційна робота відповідає завданню. Текст роботи послідовний та зрозумілий, оформлення пояснювальної записки та демонстраційних слайдів якісне.

До недоліків роботи слід віднести:

- доцільно було б для дослідження використати трафік декількох видів DDoS-атак;
- при виборі вихідних даних доцільно було б збільшити кількість Dataset.

Проте, зазначені недоліки не знижують цінності виконаної роботи.

У цілому, кваліфікаційна робота здобувача Севрюкова О.В. відповідає вимогам до випускних кваліфікаційних робіт здобувачів другого (магістерського) рівня та заслуговує оцінки «відмінно».

Здобувач Севрюкова О.В. заслуговує присвоєння кваліфікації магістр з кібербезпеки за заявленою спеціальністю 125 «Кібербезпека».

Рецензент,
завідувач кафедри
інформаційних технологій,
к.т.н., доцент

Т.І. Григор'єва

UNICHECK
by Turnitin

Ім'я користувача:
Анна Серединко

ID перевірки:
1016014317

Дата перевірки:
17.12.2023 18:59:46 MSK

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
17.12.2023 19:04:44 MSK

ID користувача:
100001433

Назва документа: MP_Севрюков_О_2023_16_12_2023_после_плагиата

Кількість сторінок: 66 Кількість слів: 11729 Кількість символів: 88442 Розмір файлу: 6.73 MB ID файлу: 1015700948

25.6%
Схожість

Найбільша схожість: 8.91% з джерелом з Бібліотеки (ID файлу: 1015699644)

19.6% Джерела з Інтернету	890	Сторінка 68
10.5% Джерела з Бібліотеки	88	Сторінка 71

0% Цитат

Вилучення цитат вимкнене

Вилучення списку бібліографічних посилань вимкнене

0%
Вилучень

Немає вилучених джерел

Модифікації

Виявлено модифікації тексту. Детальна інформація доступна в онлайн-звіті.

Замінені символи	28

РЕФЕРАТ

Текстова частина магістерської роботи: 70 с., 16 рис., 14 табл.,18 джерел.

МЕРЕЖНА АТАКА, DDoS-АТАКА, ТРАФІК, АНОМАЛІЯ, HTTP FLOOD, ICMP FLOOD, SYN FLOOD, UDP FLOOD, MAC FLOOD, UDP FLOOD, PING OF DEATH, SLOWLORIS, МАШИННЕ НАВЧАННЯ, МЕТОД ЛОГІСТИЧНОЇ РЕГРЕСІЇ, МЕТОД К-НАЙБЛИЖЧИХ СУСІДІВ, МЕТОД ВИПАДКОВОГО ЛІСУ, КЛАСИФІКАЦІЯ

Об'єкт дослідження – мережний трафік DDoS-атак.

Мета дослідження – аналіз аномалій мережного трафіку DDoS-атак з метою підвищення точності їхнього детектування.

Методи дослідження – статистичні методи, методи машинного навчання, метод логістичної решресії, метод k-найближчих сусідів, метод випадкового лісу.

У магістерській роботі проведено дослідження аномалій мережного трафіка DDoS-атак. Проаналізовано основні види DDoS-атак та основні характеристики трафіку атак. Запропоновано дослідження аномалій мережного трафіка DDoS-атак на базі статистичних методів та алгоритмів класифікації машинного навчання, таких як, алгоритм логістичної регресії, алгоритм k-найближчих сусідів та алгоритм випадкового лісу. Результати дослідження дозволять забезпечити необхідну точність визначення DDoS-атак.

ABSTRACT

Textual part of the master's thesis: 70 p., 16 figures, 14 tables, 18 sources.

NETWORK ATTACK, DDoS ATTACK, TRAFFIC, ANOMALY, HTTP FLOOD, ICMP FLOOD, SYN FLOOD, UDP FLOOD, MAC FLOOD, UDP FLOOD, PING OF DEATH, SLOWLORIS, MACHINE LEARNING, LOGISTIC REGRESSION METHOD, K-NEAREST NEIGHBOR METHOD, RANDOM FOREST METHOD, CLASSIFICATION

The object of the study is the network traffic of DDoS attacks.

The purpose of the study is to analyze the anomalies of network traffic of DDoS attacks in order to increase the accuracy of their detection.

The research methods are statistical methods, machine learning methods, logistic solution method, k-nearest neighbor method, random forest method.

In the master's thesis, the study of anomalies of network traffic of DDoS attacks is carried out. The main types of DDoS attacks and the main characteristics of attack traffic are analyzed. A study of anomalies of network traffic of DDoS attacks on the basis of statistical methods and algorithms for the classification of machine learning, such as the algorithm of logistic regregation, the algorithm of k-nearest neighbors and the algorithm of a random forest is proposed. The results of the study will ensure the necessary accuracy of determination DDoS attacks.

ЗМІСТ

ВСТУП

Сучасний технологічний розвиток інфокомунікаційних систем та мереж супроводжується необхідністю забезпечення концепції безпеки мережевої інфраструктури від різних атак, в тому числі атак, що призводять до відмови в обслуговуванні DDoS-атак (SYN-Flood, ICMP-Flood, UDP-Flood). Основною особливістю таких атак є порівняно низька інтенсивність трафіку з атаками на мережевому рівні. При цьому DDoS-атаки виявляються досить важко через значну кількість джерел атак і особливостей характеристик трафіку адже характеристики шкідливого трафіку DDoS-атак іноді можуть бути майже невідмінними від легітимного.

Трафік, який генерується в мережі DDoS-атаками, досить різнорідний, іноді – це низькошвидкісна передача даних з невеликою інтенсивністю, а іноді, передача високошвидкісна передача трафіку атаки зі значною інтенсивністю. Це визначає характеристики трафіку, які іноді мають значні і часті сплески інтенсивності, а іноді, майже рівномірну структуру. При цьому, при DDoS-атаці може генеруватися трафік різних характеристик для різних потоків та структур.

Класичні методи виявлення DDoS-атак на відмову в обслуговуванні базуються на детектуванні аномалій трафіку, які характерні для цього виду атак. В той же час DDoS-атаки на відмову в обслуговуванні прикладного рівня OSI (Open Systems Interconnection) не вимагають генерації значного обсягу трафіку, тому атаки такого типу досить складно ідентифікуються звичайними системами. Сучасні заходи для захисту від DDoS-атак часто є неефективними, що викликає значне зростання їхньої кількості.

В цьому змісті й розрахована магістерська робота, яка розглядає можливості дослідження характеристик трафіку DDoS-атак та в подальшому, базуючись на цих результатах, проведення аналізу такого трафіку. Рішення цієї задачі дозволить накдати низку рекомендаційй щодо визначення різних видів аномалій, спричинених DDoS-атакою.

1 ОГЛЯД ТЕХНОЛОГІЙ ТА МЕТОДІВ ВИЯВЛЕННЯ ТА ПОПЕРЕДЖЕННЯ АНОМАЛІЙ МЕРЕЖНОГО ТРАФІКУ DDoS-АТАК

Одним із найбільш розповсюджених видів мережевих атак є атаки типу «відмова в обслуговуванні» – DDoS (Distributed Denial of Service Attack). DDoS (Distributed Denial of Service Attack) – це атака, метою якої є викликати відмову в обслуговуванні на боці сервера шляхом його перевантаження з подальшою недоступністю. Характерною особливістю таких атак є їх виконання з великої кількості комп'ютерів, тому вони мають характер розподіленого мережного вторгнення. В результаті DDoS-атак виникають переповнення смуги пропускання вузла, що атакується, тому доступ до веб-ресурсу легітимних користувачів сттає істотно утрудненим. Відомо, що ці атаки можуть призвести до повного блокування системи або веб-ресурсу [1-2].

DDoS-атаки можуть відбуватися на різних рівнях мережевої інфраструктури. Такі атаки можуть вражати всі компоненти мережі: канали передачі, атаки на брандмауер, на вбудовані послуги та веб-ресурси.

Сьогодні DDoS-атаки стали багатовекторними. Якщо ще нещодавно атака складалася з одного на систему, то тепер атака спрямована на переповнення як мережевих ресурсів, так й обчислювальних ресурсів сервера одночасно, і захиститися від неї стає значно важче. Різко зросла кількість атак на прикладному рівні мережі, спрямованих на веб-ресурси. Відомо, що крім складності DDoS-атак вони стали також більш тривалими за часом [1-3].

Виявлення DDoS-атак є надзвичайно складним та трудомістким завданням. Існуючі системи виявлення вторгнень вимагають постійного оновлення набору правил для завчасного виявлення загроз. Крім того, з кожним роком пропускна спроможність мережних каналів зростає, тому застосування лише апаратних засобів детектування мережного трафіку є недоцільним. При цьому використання класичних засобів захисту, таких як, наприклад, міжмережні екрани на сьогоднішній день є нефективним через використання певного набору правил, відповідно до якого здійснюється фільтрація всіх даних. Тим не менш, навіть періодичного оновлення набору правил може бути недостатньо для того, щоб система завжди залишалася в актуальному стані [3].

Найчастіше для детектування таких атак проводиться аналіз аномалій мережного трафіку, тобто за штатних умов роботи мережі ведеться пошук відхилень від контрольних характеристик мережного трафіку. Таким чином, з

підвищенням кількості та складності мережевих розподілених DDoS-атак необхідно підвищення точності та швидкості детектування мережевих атак.

1.1 Аномалії мережного трафіку

Аномалія мережного трафіку – це відхилення характеристик трафіку від штатного. Природа самих аномалій є різна і може виявлятися при технологічних несправностях, помилках користувачів або навмисних протиправних дій проти системи, наприклад, зломах [3].

Існує безліч методів детектування різних аномалій, водночас далеко не всі методи є універсальними. Аномалії поділяються на три категорії [3]:

– аномалії, при яких одиничний екземпляр даних може розглядатися як аномалія;

– друга категорія – умовна, у яких аномалія екземпляра розглядається лише тому випадку, коли виявляє себе за певних умов чи певному контексті, інакше кажучи при аномалія вважається такий лише певних обстановках;

– третій вид аномалій – колективний, інакше кажучи, даному виду аномалій відповідає ціла послідовність примірників, яка згодом і формує колективний ланцюжок аномалій (наприклад, часовий ряд).

Очевидно, характер поведінки шкідливого мережевого трафіку щодо інших відповідає першому виду аномалій. Записана сесія мережного трафіку, який було виявлено як аномальний, є архівними даними. Завдання підбору методу має ґрунтуватися, в першу чергу, на оцінці ступеня ймовірності того, що аналізований екземпляр буде аномальним по відношенню до іншого вхідного мережного трафіку.

Режим розпізнавання аномалій у мережевому трафіку, для визначення раніше невідомих атак, є режимом розпізнавання без вчителя, іншими словами, виноситься припущення про те, що такі аномальні сплески буде зустрічатися рідко. Цей метод не передбачає роботи з потоковими даними, оскільки вимагає обробки всього набору даних, необхідного для аналізу.

Залежно від поставленого завдання визначення аномалій поділяються на п'ять груп [3]:

1. Методи, засновані на вирішенні задачі класифікації. Як нормальна поведінка призначається один або кілька екземплярів цільової функції. Примірник, який не відповідає жодному класу, буде аномальним по відношенню

до інших екземплярів у аналізованому наборі даних. Цей тип відноситься до машинного навчання з учителем.

2. Методи, засновані на задачі кластеризації. Завдання кластеризації не потребує наявності цільової функції. Нормальні екземпляри класів утворюватимуть значно більшу щільність, ніж аномальні кластери. На жаль, стандартні алгоритми кластеризації мають високий рівень помилкових спрацьовувань через розмитість кордону між нормальними і аномальними примірниками.

3. Методи, що ґрунтуються на статистичному аналізі даних. У разі будується нормальна модель поведінки системи, будь-яке відхилення від поведінки вважатиметься аномальним. Принципова проблема полягає в тому, що якщо природа аномалій невідома заздалегідь, тоді буде складно визначити точність статистичного розподілу та порога.

4. Метод найближчого сусіда. Цей метод заснований на Евклідовому відстані між екземплярами. Насамперед, необхідно зрозуміти міру схожості аналізованих екземплярів. Обчислюється відстань до найближчого екземпляра. Коли екземпляр віддалений від сусіда – даний примірник позначається викидом і позначається аномальним.

5. Методи спектрального аналізу. Даний вид методу характеризується завданням апроксимації вхідних атрибутів даних. Найчастіше для аналізу мережевого трафіку будується часовий ряд, щоб отримати частотні показники. Згодом цей аналіз порівнюється зі спектральним аналізом ряду, отриманим під час атаки. Наявність відхилень свідчить про наявність атаки. В той же час спрогнозувати даним методом раптову активність користувачів не є можливим, тому для DDoS атак спектральний аналіз є ефективним лише для аналізу гістограми вхідного потоку трафіку, а не для детектування аномальних екземплярів. Тим самим, велика ймовірність помилкового спрацьовування методу на трафіку, що не містить DDoS атак.

1.2 DDoS-атака

DDoS-атака – це атака, метою якої є викликати відмову в обслуговуванні на боці сервера шляхом його перевантаження з подальшою недоступністю. Виконується шляхом відправки величезної кількості хаотичних шкідливих запитів з різних IP-адрес, використовуючи вразливості та недоліки безпеки (рис. 1). У висновку, трафік і навантаження постійно зростають, а ресурси і ліміти

пропускної здатності сервера вичерпуються. Сервер, не взмозі обробити велику кількість запитів одночасно, віддає код відповіді 5xx – відмова в обслуговуванні. Різниця DDOS і DOS полягає у тому, що DOS-атака виконується з одного джерела [3].

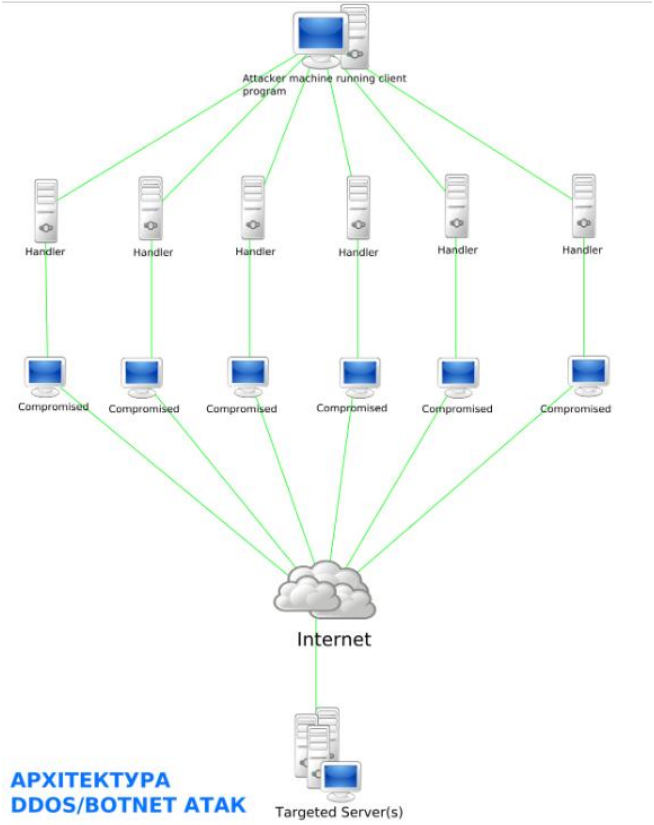


Рисунок 1.1 – DDoS-атака

2 ДЕТЕКТУВАННЯ АНОМАЛІЙ МЕРЕЖНОГО ТРАФІКУ DDoS-АТАК ЗА ДОПОМОГОЮ СТАТИСТИЧНИХ МЕТОДІВ

Одним з ефективних рішень запобігання вторгненням в умовах їх постійного ускладнення та кількісного зростання є засоби моніторингу та багатофункціонального аналізу трафіку. Моніторинг мережного трафіку сьогодні націлений насамперед на оперативне виявлення «сплесків» мережевого трафіку інфокомунікаційної мережі, що є результатом несанкціонованого доступу, вірусів, атак, вторгнень та інших загроз інформаційній безпеці [4-5].

Серед таких загроз найчастіше зустрічаються атаки типу DDoS (Distributed Denial of Service), які стають все більш інтенсивними та руйнівними, а їхня кількість безперервно зростає. При цьому виявляються вони досить важко через значну кількість джерел атак і особливостей трафіку. Характеристики шкідливого трафіку атак іноді можуть бути майже невідмінними від легітимного [4-5].

Процес виявлення кібератаки в мережі найчастіше заснований на порівнянні характеристик трафіку на незначному відрізку часу вторгнення з відповідною легітимною трасою трафіку, розглянутої за тривалий період часу. У цьому випадку, якщо виявлено значні «сплески» трафіку, виникає необхідність знайти такий метод виявлення кібератак, у якому похибка виявлення буде мінімальною.

Для протидії зловмисникам сьогодні створено низку складних систем виявлення вторгнень IDS/IPS (Intrusion detection system/Intrusion prevention system), міжмережеві екрани, аналізатори трафіку на базі утиліт та сніферів (tcpdump, Wireshark, Snort), протоколи SNMP (NetFlow). Для оцінки трафіку ці системи використовують різні методи, серед яких: статистичні методи, інтелектуальний аналіз даних, штучний інтелект, вейвлет-аналіз та інші [4-5].

2.1 Основні види DDoS-трафіку

Найпростіший вид DDoS-трафіку – це HTTP-запити [3]. За допомогою таких запитів, наприклад будь-який відвідувач спілкується з вашим сайтом за допомогою браузера. В основі запиту лежить HTTP-заголовок. HTTP заголовки - це поля, які описують, який саме ресурс запитується, наприклад, URL-адреса або форма, або JPEG. Також заголовки HTTP інформують веб-сервер, який тип браузера використовується. Найбільш поширені HTTP заголовки: ACCEPT, LANGUAGE та USER AGENT [3].

Сторона, що запитує, може використовувати скільки завгодно заголовків, надаючи їм потрібні властивості. Зловмисники, які проводять DDoS-атаку, можуть змінювати ці та багато інших HTTP-заголовків, роблячи їх важкорозпізнаними для виявлення атаки. Крім того, HTTP заголовки можуть бути написані таким чином, щоб управляти кешуванням і проксі-сервісами. Наприклад, команда проксі-серверу не кешувати інформацію [3].

HTTP GET-запит – це метод, який запитує інформацію на сервері. Цей запит може попросити сервера передати якийсь файл, зображення, сторінку або скрипт, щоб відобразити їх у браузері.

HTTP(S) GET-флуд - метод DDoS атаки прикладного рівня (7) моделі OSI, при якому атакуючий посилає потужний потік запитів на сервер з метою переповнення його ресурсів. У результаті сервер не може відповідати не тільки на запити хакерів, але і на запити реальних клієнтів.

HTTP POST-запит — метод, коли дані містяться в тіло запиту для подальшої обробки на сервері. HTTP POST-запит кодує інформацію, що передається, і поміщає на форму, а потім відправляє цей контент на сервер. Цей метод використовується за необхідності передавати великі обсяги інформації чи файли.

HTTP(S) POST-флуд - це тип DDoS-атаки, при якому кількість POST-запитів переповнюють сервер так, що сервер не в змозі відповісти на всі запити. Це може призвести до виключно високого використання системних ресурсів і, згодом, до аварійної зупинки сервера.

Кожен із описаних вище HTTP-запитів може передаватися за захищеним протоколом HTTPS. У цьому випадку всі дані, що пересилаються між клієнтом (зловмисником) і сервером, шифруються. Виходить, що «захищеність» тут грає на руку зловмисникам: щоб виявити зловмисний запит сервер повинен спочатку розшифрувати його. Тобто. розшифровувати доводиться весь потік запитів, яких під час DDoS-атаки надходить дуже багато. Це створює додаткове навантаження на сервер-жертву.

SYN-флуд (TCP/SYN) встановлює напіввідкриті з'єднання з вузлом. Коли жертва приймає SYN-пакет через відкритий порт, вона повинна надіслати у відповідь пакет SYN-ACK і встановити з'єднання. Після цього ініціатор надсилає одержувачу відповідь з ACK-пакетом. Цей процес умовно називається рукостисканням. Однак, під час атаки SYN-флудом рукостискання не може бути завершено, оскільки зловмисник не відповідає на SYN-ACK сервера-жертви. Такі

з'єднання залишаються напіввідкритими до закінчення тайм-ауту, черга на підключення переповнюється і нові клієнти не можуть підключитися до сервера.

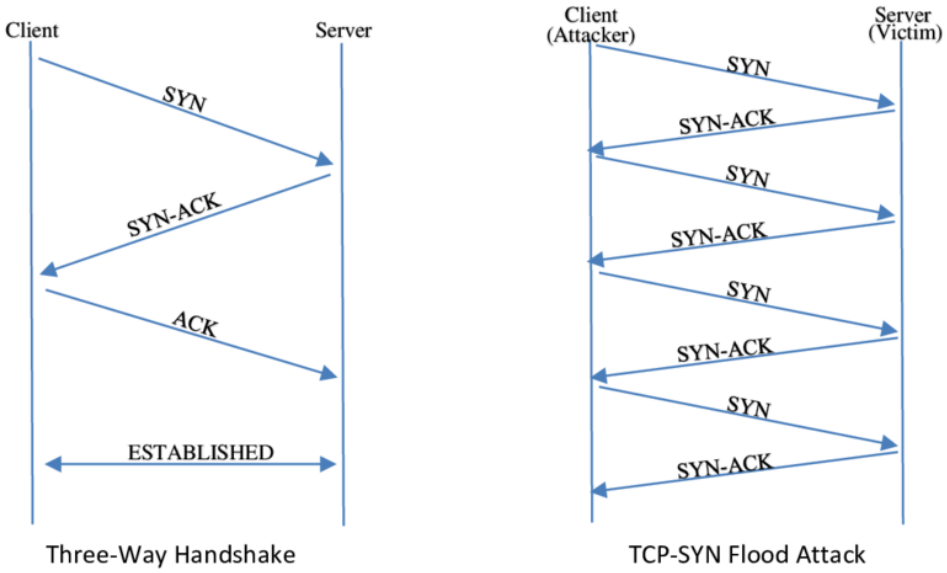


Рисунок 2.1 – Приклад SYN-флуд

UDP-флуд найчастіше використовуються для широкосмугових DDoS-атак через їхню безсеансовість, а також простоту створення повідомлень UDP різними мовами програмування. UDP-флуд – працює поверх протоколу IP. Як такої установки з'єднання не відбувається - дані надсилаються без контролю цілісності. У результаті зловмисник отримує можливість підмінити IP-адресу джерела, а потім розсилати пакети зі свого пристрою так, щоб виглядало, що вони приходять з різних місць. У такому вигляді вони приходять на сервер, підвищуючи час очікування відповіді (рис. 2.2) [3].

Фрагментований UDP-флуд – різновид UDP-флуду, але має ще одну додаткову особливість. На сервер компанії-жертви приходить пакет із зазначенням, що це лише його частина. Сервер резервує ресурс для збирання, а нові фрагменти так і не надходять.

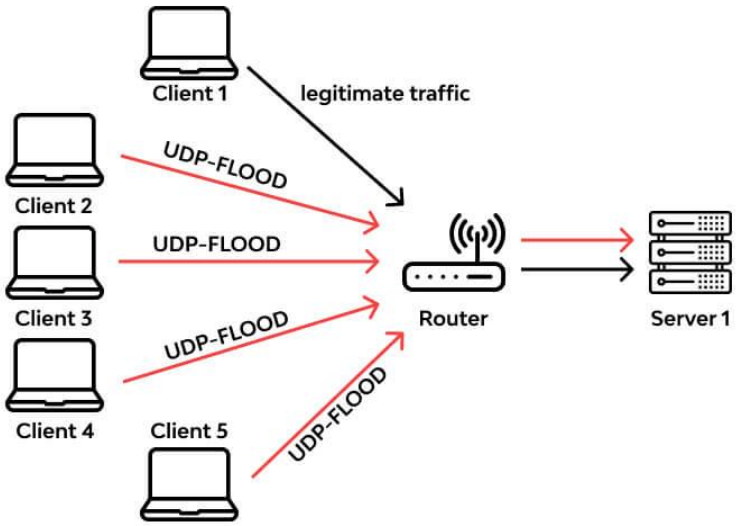


Рисунок 2.2 – Приклад UDP-флуд

Smurf-атака - при такому способі сервер атакованої компанії заповнюється підробленими пакетами вихідних IP-адрес та пакетами ICMP. Цей різновид атаки отримав назву від шкідливої утиліти DDoS. Невеликий ICMP-пакет, що генерується шкідливою утилітою, може сильно нашкодити IT-системі. Тому цей вид атаки назвали Smurf (рис. 2.3) [3].

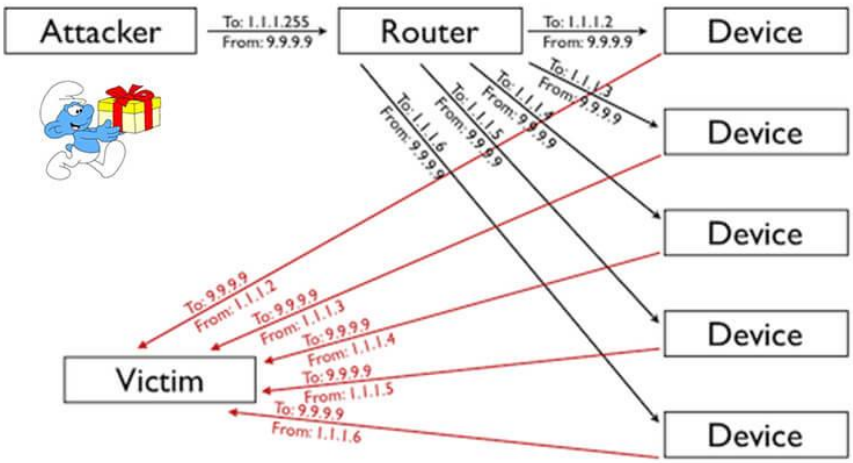


Рисунок 2.3 – Приклад Smurf-атаки

ICMP-флуд. Протокол міжмережевих керуючих повідомлень ICMP використовується в першу чергу для передачі повідомлень про помилки і не використовується для передачі даних. ICMP-пакети можуть супроводжувати TCP-пакети під час з'єднання з сервером. ICMP-флуд - метод DDoS атаки на 3-му рівні

моделі OSI, що використовує ICMP-повідомлення для перевантаження мережного атакованого каналу [3].

MAC-флуд - рідкісний вид атаки, при якому атакуючий посилає множинні порожні Ethernet-фрейми з різними MAC-адресами. Мережеві свитчі розглядають кожну MAC-адресу окремо і, як наслідок, резервують ресурси під кожну з них. Коли вся пам'ять на світчі використана, він або перестає відповідати, або вимикається. На деяких типах роутерів атака MAC-флудом може спричинити видалення цілих таблиць маршрутизації, таким чином порушуючи роботу цілої мережі [3].

2.2 Класифікація DDoS-атак за механізмом дії

Перша група – різні види флуду. Мета флуду - створити потужний потік запитів (пакетів даних), щоб він забив виділену жертві смугу і перекрив весь трафік. До цієї групи належать: DNS-ампліфікація, фрагментований UDP-флуд, ICMP-флуд, NTP-флуд, NTP-ампліфікація, фрагментований ACK-флуд, Ping-флуд, UDP-флуд, UDP-флуд за допомогою ботнета, VoIP-флуд, флуд медіаданими, атака широкомовними ICMP ECHO-пакетами, атака широкомовними UDP-пакетами, фрагментований ICMP-флуд, DNS-флуд та інші атаки з посиленням [3-5].

Друга група - атаки, які використовують уразливості стека мережевих протоколів: SYN-флуд (SYN Flood), IP-null-атака, атака підробленими TCP-сесіями, TCP-null-атака, атаки з модифікацією поля TOS, ACK/PUSH ACK-флуд , SYN-ACK-флуд, RST/FIN-флуд, TCP null/IP-null-атака, атака підробленими TCP-сесіями з кількома SYN-ACK, атака з заміною адреси відправника адресою одержувача, атака за допомогою перенаправлення трафіку високонавантажених сервісів; Ping смерті, атака підробленими TCP-сесіями з кількома ACK [3].

Третя група - DDoS-атаки на прикладний рівень або рівень додатків: HTTP-флуд, атаки з метою відмови програми, HTTP-флуд одиночними запитами, атака фрагментованими HTTP-пакетами, HTTP-флуд одиночними сесіями, сесійні атаки, атака повільними сесіями [3-5].

2.3 Класифікація DDoS-атак за рівнями OSI

DDoS-атаки можливі на кажному із семи рівнів моделі OSI (рис. 2.4). Розкглянемо кожний з рівнів детально. Характеристики DDoS-атаки на сьомому, прикладному рівні OSI показані в табл. 2.1 [3-5].

Таблиця 2.1 – Характеристики DDoS-атак на прикладному рівні моделі OSI

Типи даних	Дані
Опис рівня	Початок створення пакетів даних. Підєднання та доступ до даних. Користувацькі протоколи FTP, SMTP, Telnet, RAS
Протоколи	FTP, HTTP, POP3, SMTP та шлюзи, які їх використовують
Приклади DDoS-технологій	PDF GET запити, HTTP GET, HTTP POST (форми веб-сайтів: логін, завантаження фото/відео, підтвердження зворотнього зв’язку)
Наслідки DDoS-атаки	Нестача ресурсів. Надмірне споживання системних ресурсів службами на сервері, що атакується.

DDoS-атаки сьомого, прикладного рівня OSI виявляються за допомогою моніторингу додатків, систематичного моніторингу програмного забезпечення, який використовує певний набір алгоритмів, технологій та підходів (залежно від платформи, на якому це програмне забезпечення використовується) для виявлення уразливостей програм (атаки 7 рівня). Ідентифікувавши такі атаки, їх можна раз і назавжди зупинити та відстежити їхнє джерело. На цьому шарі це здійснюється найпростіше.

Характеристики DDoS-атаки на шостому, представницькому рівні OSI показані в табл. 2.2.

Таблиця 2.2 – Характеристики DDoS-атак на представницькому рівні моделі OSI

Типи даних	Дані
Опис рівня	Трансляція даних від джерела до отримувача
Протоколи	Протоколи стиснення та кодування даних (ASCII, EBCDIC)
Приклади DDoS-технологій	Підроблені SSL запити: перевірка шифрованих пакетів SSL дуже ресурсомістка, зловмисники використовують SSL для HTTP-атак на сервер жертви
Наслідки DDoS-атаки	Атаковані системи можуть перестати приймати SSL з'єднання або автоматично перевантажуватися

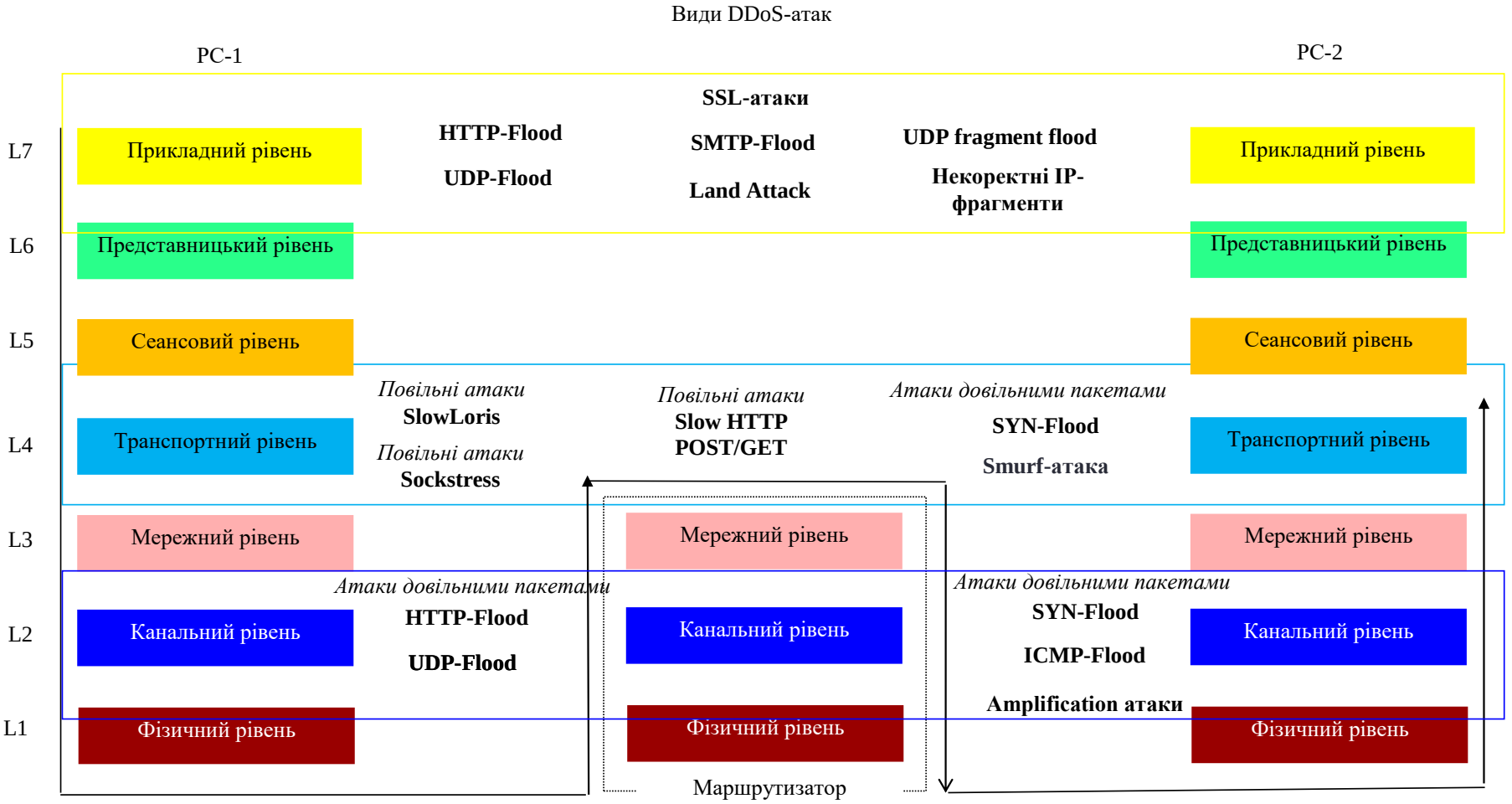


Рисунок 2.4 – Види DDoS-атак [5]

DDoS-атаки шостого, представницького рівня OSI запобігаються за допомогою розподілу шифруючої SSL інфраструктури (тобто розміщення SSL на відмінному сервері, якщо це можливо) та перевірки трафіку додатків на предмет атак або порушення політик на платформі додатків. Платформа гарантує, що трафік шифрується та відправляється назад початковій інфраструктурі з розшифрованим контентом, що знаходився у захищеній пам'яті безпечного вузла.

Характеристики DDoS-атаки на п’ятому, сеансовому рівні OSI показані в табл. 2.3 [3].

Таблиця 2.3 – Характеристики DDoS-атак на сеансовому рівні моделі OSI

Типи даних	Дані
Опис рівня	Керування установкою та завершенням з'єднання, синхронізацією сеансів зв'язку в рамках операційної системи через мережу (наприклад, коли ви виконуєте вхід/вихід)
Протоколи	Протоколи входу/виходу (RPC, PAP)
Приклади DDoS-технологій	Атака на протокол Telnet використовує слабкі місця програмного забезпечення Telnet-сервера на світчі, роблячи сервер недоступним
Наслідки DDoS-атаки	Унеможливлює адміністратора управління свитчем

DDoS-атаки на п’ятому, сеансовому рівні OSI блокуються за допомогою підтримки прошивки апаратного забезпечення в актуальному стані, що дозволяє зменшити ризики появи загрози.

Характеристики DDoS-атаки на четвертому, транспортному рівні OSI показані в табл. 2.4 [3].

Таблиця 2.4 – Характеристики DDoS-атак на транспортному рівні моделі OSI

Типи даних	Сегменти
Опис рівня	Забезпечення безпомилкової передачі між вузлами, управління передачею повідомлень з 1 по 3 рівень
Протоколи	Протоколи TCP, UDP
Приклади DDoS-технологій	SYN-флуд, Smurf-атака (атака ICMP-запитами зі зміненими адресами)

Наслідки DDoS-атаки	Досягнення меж за шириною каналу чи кількістю допустимих підключень, порушення роботи мережного обладнання

DDoS-атаки четвертого, транспортного рівня OSI запобігаються за допомогою фільтрації DDoS-трафіку, відомої як blackholing - методу, який часто використовується провайдерами для захисту клієнтів. Однак, цей підхід робить сайт клієнта недоступним як для трафіку зловмисника, так і для легального трафіку користувачів. Проте блокування доступу використовується провайдерами у боротьбі з DDoS-атаками для захисту клієнтів від таких загроз, як уповільнення роботи мережного обладнання та відмова роботи сервісів.

Характеристики DDoS-атаки на третьому, мережному рівні OSI показані в табл. 2.6.

Таблиця 2.6 – Характеристики DDoS-атак на мережному рівні моделі OSI

Типи даних	Пакети
Опис рівня	Маршрутизація та передача інформації між різними мережами
Протоколи	Протоколи IP, ICMP, ARP, RIP та роутери, які їх використовують
Приклади DDoS-технологій	ICMP-флуд — DDos-атаки на третьому рівні моделі OSI, які використовують ICMP-повідомлення для перевантаження пропускної спроможності цільової мережі
Наслідки DDoS-атаки	Зниження пропускної спроможності атакованої мережі та можлива перевантаженість брандмауера

Для запобігання DDoS-атак третього, мережного рівня OSI доцільно обмежити кількість оброблених запитів за протоколом ICMP та скоротити можливий вплив цього трафіку на швидкість роботи Firewall та пропускну спроможність інтернет-смуги.

Характеристики DDoS-атаки на другому, канальному рівні OSI показані в табл. 2.7 [3].

Таблиця 2.7 – Характеристики DDoS-атак на канальному рівні моделі OSI

Типи даних	Кадри
Опис рівня	Встановлення та супровід передачі повідомлень фізично
Протоколи	Протоколи 802.3, 802.5, а також контролери, точки доступу та мости, які їх використовують
Приклади DDoS-технологій	MAC-флуд – переповнення пакетами даних мережних комутаторів
Наслідки DDoS-атаки	Потоки даних від відправника одержувачу блокують роботу всіх портів

Для запобігання DDoS-атак другого, канального рівня OSI доцільно встановити на комутаторах таким чином, що кількість MAC адрес обмежується надійними, які проходять перевірку автентифікації, авторизації та обліку на сервері (протокол AAA) і фільтруються.

Характеристики DDoS-атаки на першому, фізичному рівні OSI показані в табл. 2.8.

Таблиця 2.8 – Характеристики DDoS-атак на фізичному рівні моделі OSI

Типи даних	Біти
Опис рівня	Передача двійкових даних
Протоколи	Протоколи 100BaseT, 1000 Base-X, а також концентратори, розетки та патч-панелі, які їх використовують
Приклади DDoS-технологій	Фізична руйнація, фізична перешкода роботі чи управлінню фізичними мережевими активами
Наслідки DDoS-атаки	Мережеве обладнання стає непридатним і вимагає ремонту для відновлення роботи

Для запобігання DDoS-атак другого, канального рівня OSI доцільно використовувати систематичний підхід до моніторингу роботи фізичного мережного обладнання [3].

Хоча атака можлива на будь-якому рівні, особливої уваги слід приділяти атакам на 3-4 і 7 рівнях моделі OSI.

DDoS-атаки на 3-му та 4-му рівні - інфраструктурні атаки - типи атак, засновані на використанні великого обсягу, потужного потоку даних (флуд) на

рівні інфраструктури мережі та транспортному рівні з метою уповільнити роботу веб-сервера, «заповнити» канал , і зрештою завадити доступу інших користувачів до ресурсу. Ці типи атак зазвичай включають ICMP-, SYN- і UDP-флуд [3].

DDoS атака на 7-му рівні - атака, що полягає в навантаженні деяких специфічних елементів інфраструктури сервера додатків. Атаки 7-го рівня особливо складні, приховані та важкі для виявлення в силу їхньої подібності з корисним веб-трафіком. Навіть найпростіші атаки 7-го рівня, наприклад, спроба входу в систему під довільним ім'ям користувача і паролем або довільний пошук, що повторюється, на динамічних веб-сторінках, можуть критично завантажити CPU і бази даних. Також DDoS зловмисники можуть неодноразово змінювати сигнатури атак 7-го рівня, роблячи їх ще складнішими для розпізнавання та усунення [3].

Основний підхід щодо запобігання DDoS-атакам наведено в табл. 2.9.

Таблиця 2.9 – Підхід щодо запобігання DDoS-атакам

Обладнання	Рівень	Оптимізація	DDoS-захист
Брандмауер	4-7	Перевірка потоку, глибинна перевірка	Екрани, обмеження сеансу SYN cookie
Роутер	3-4	Пакетна перевірка, фреймова перевірка	Лінійні списки контролю доступу, обмеження швидкості

До запобігання атак можуть бути використані наступне [3]:

– використання брандмауерів з динамічною перевіркою пакетів;

– використання динамічних механізмів SYN проксі;

– обмеження кількості SYN-ів за секунду для кожної IP-адреси;

– обмеження кількості SYN-ів за секунду для кожної віддаленої IP-адреси;

– встановлення екранів ICMP флуду на брандмауері;

– встановлення екранів UDP флуду на брандмауері;

– обмеження швидкості роутерів, що примикають до брандмауерів та мережі.

2.4 Характеристики DDoS-атак

Ключові характеристики атак DDoS типу «відмова в обслуговуванні» – потужність, частота та тривалість. Саме потужність є найважливішим фактором для атакованої інфраструктури — адже атаки, які переповнять доступні канали зв'язку, виведуть систему з ладу. Потужність DDoS-атаки визначається на підставі кількох параметрів [3]:

- Обсяг - вимірюється в бітах в секунду BPS,
- Швидкість – вимірюється в пакетах за секунду PPS
- Кількість запитів на секунду – RPS.

Розглянемо кожен із цих показників докладніше.

Обяг DDoS-атаки, зазвичай вимірюється в бітах за секунду bps, у мегабітах за секунду Mbps, а останні роки часто й у сотнях гігабіт Gbps.

У 2022 році найбільше зростання атак в діапазоні 500-Мбіт/с - 1 Гбіт/с, проте великі атаки в 100 Гбіт/с і більше стали відбуватися частіше. Згідно зі статистикою NETSCOUT за січень 2023 року, за обсягом більшість DDoS-атак знаходилися в діапазонах 100 Mbps – 1 Gbps та 1 Gbps – 10 Gbps [3].

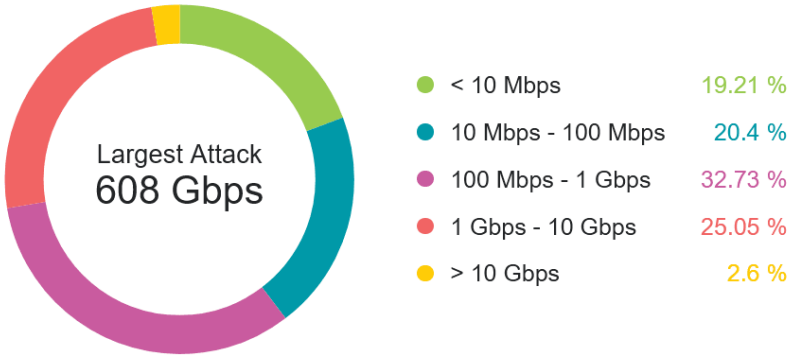


Рисунок 2.3 – Розподіл DDoS-атак за обсягом за січень 2023 року (NETSCOUT Omnis Threat Horizon)

Об'ємні атаки націлені на те, щоб переповнити смуги пропускання. До цієї категорії входить ряд різних видів флудів - UDP, ICMP та інші потоки сфальшованих пакетів.

Наступна характеристика – швидкість атаки позначає, скільки пакетів в секунду pps передає організатор DDoS-атаки. У таких одиницях вимірюють атаки мережного рівня. Чим більше пакетів, що атакують, передається в момент, тим швидше перевантажуються канали зв'язку, і тим більше шансів, що сайт вийде з ладу. Більшість атак мають порівняно невисоку швидкість, один мільйон пакетів

за секунду і менше, проте для сайтів без захисту і цього може бути цілком достатньо, щоб вийти з ладу.

Згідно зі статистикою NETSCOUT за січень 2023 року, за швидкістю більшість DDoS-атак знаходилися в діапазонах 10 kpps – 100 kpps та 100 kpps – 1 Mpps.

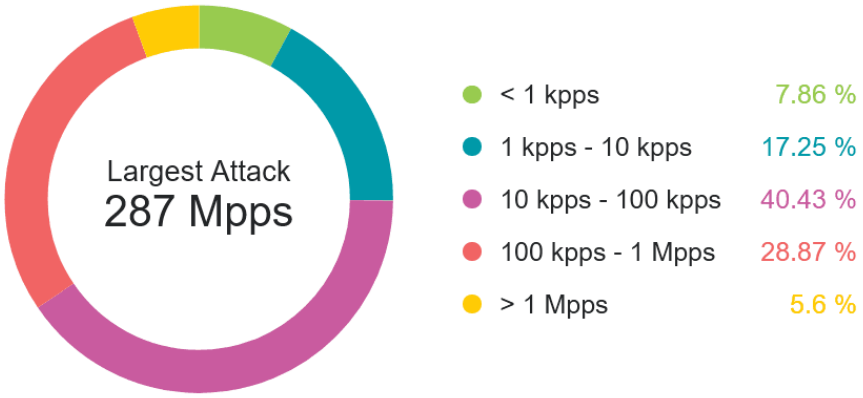


Рисунок 2.4 – Розподіл DDoS-атак за січень 2023 року (NETSCOUT Omnis Threat Horizon)

Найбільш ефективний спосіб боротьби з атаками на мережевий рівень, особливо потужними, - фільтрація вхідного трафіку на рівні дата-центру або підключення зовнішнього захисту від DDoS на рівнях L3-L4 . Шкідливі запити блокуються, а корисний трафік (справжні користувачі) може без проблем потрапити на сайт.

Наступна характеристика – кількість запитів за секунду rps визначає кількість звернень до сервера. Будь-яке обладнання має межу за кількістю оброблюваних запитів, тому їх аномальна кількість може викликати збій. У випадку рекордних кіберинцидентів кількість запитів сягала десятків мільйонів. У rps вимірюють атаки на прикладному рівні (L7 по системі OSI), які націлені на вичерпання ресурсів сервера і виведення з ладу веб-додатки через типові запити, що повторюються, до бази даних, пам'яті або диска.

Рядова DDoS-атака може вимірюватися тисячами та десятками тисяч запитів. Середня кількість запитів за секунду під час такої атаки, за даними DDoS-Guard на січень 2023 року, становила 10-20 з однієї адреси. Важливо, що для сайтів з низьким навантаженням навіть атака в 10 rps може спричинити збої в роботі, тим більше, що малопотужність атак компенсується їхньою підвищеною частотою. Під час найпотужнішої атаки, зареєстрованої DDoS-Guard у січні 2023, фіксувалося майже 60 тисяч запитів на секунду.

Налаштування параметрів на стороні сайту - наприклад, обмеження за кількістю запитів з одного IP - може бути тимчасовим рішенням, проте є ризик погіршення досвіду користувача через те, що справжніх відвідувачів система прийме за ботів.

Для боротьби з атаками на рівні програм використовується моніторинг поведінки відвідувачів сайту. Відомі патерни ботів розпізнаються і блокуються, а підозрілі відвідувачі перевіряються повторно і, за необхідності, підтверджують, що вони люди через технологію CAPTCHA.

Згідно проведеного аналізу, у відповідності до вищезазначеного можливо зробити наступні висновки щодо характеристик DDoS-атак:

1. Об'єктивно визначити потужності атаки може бути складно. Наприклад, якщо гранична ємність мережі 100 Гбіт/с та обладнання вийшло з ладу - відразу складно визначити, чи це була атака об'ємом 150 Гбіт/с або 1 Тбіт/с. Виміряти об'єм атаки можна по висхідних потоках (upstream) атакованої мережі. Якщо частина атакуючого трафіку була відфільтрована у процесі, розрахунок точного розміру атаки може бути складнішим. До зафіксованого обсягу потрібно додати інші характеристики атаки — швидкість і кількість запитів.

2. При захисті від DDoS велике значення має потужність атаки. Тому необхідним є вимір характеристик атакуючого трафіку, таких як обсяг, швидкість та кількість. Якщо атаки поки не надто потужні, можливо, буде достатньо автономної системи фільтрації. Важливою є саме сукупність усіх показників та процес фільтрування трафіку, що надходить, відокремлюючи шкідливі запити від легітимних.

2.5 Огляд рішень щодо запобігання DDoS-атак та аномалій мережевого трафіку DDoS-атак

Одним з ефективних рішень запобігання вторгненням в умовах їх постійного ускладнення та кількісного зростання є засоби моніторингу та багатофункціонального аналізу трафіку [4-5].

Аналізатори трафіку (spoofing) це програми для перехвату та аналізу мережевого трафіку. Сніфер може аналізувати тільки той трафік, що проход.ть безпосередньо через його мережеву карту. Є декілька способів перехвату трафіку: звичайне прослуховування мережевого інтерфейса, підключення сніферу в розрив каналу, відгалуження, аналіз побічних електромагнітних випромінювань, через так звані атаки MAC-spoofing та IP-spoofmg. IP-spoofmg - це тип зловмисної атаки, при якій зловмисник приховує справжнє джерело IP-пакетів, щоб важко

було дізнатися, звідки вони прийшли. Зловмисник створює пакети, змінюючи вихідну IP-адресу , щоб видавати себе за іншу комп'ютерну систему, приховати особу відправника або обидва. Поле заголовка підробленого пакета для вихідної IP-адреси містить адресу, яка відрізняється від фактичної IP-адреси джерела. Кінцевим користувачам важко виявити підробку IP. Ці атаки здійснюються на мережевому рівні - рівні 3 моделі зв'язку OSI. Таким чином, не буде зовнішніх ознак фальсифікації. Підроблені запити на з'єднання ззовні виглядають як законні запити на з'єднання [4-5].

Моніторинг мережного трафіку сьогодні націлений насамперед на оперативне виявлення «сплесків» мережевого трафіку інфокомунікаційної мережі, що є результатом несанкціонованого доступу, вірусів, атак, вторгнень та інших загроз інформаційній безпеці. До таких загроз відносяться атаки типу DDoS, які стають все більш інтенсивними та руйнівними, а їхня кількість безперервно зростає. При цьому виявляються вони досить важко через значну кількість джерел атак і особливостей трафіку. Характеристики шкідливого трафіку атак іноді можуть бути майже невідмінними від легітимного [4-5].

Апаратні засоби аналізу трафіку, тобто пристрої для перехоплення та\або разом з аналізом цього трафіку, прикладом є Cisco RSPAN та Kismet. Cisco RSPAN - у багатьох мережних комутаторах є можливість дзеркалювати трафік, скажімо з одного порту, на інший, або, наприклад, з VLAN зазначеного, на порт, де є аналізатор трафіку, чи якесь ПЗ. У Cisco ця технологія називається SPAN - Switch Port Analyzer та RSPAN - Remote Switch Port Analyzer.

В першу чергу, ця технологія необхідна для того, щоб переглянути трафік на якомусь порту для аналізу того, що передається в мережі. Так само вона може знадобитися, наприклад, для запису VOIP. VLAN Voice переправляє весь трафік на певний інтерфейс, а там розгорнуте деяке програмне 7 забезпечення, що має змогу записувати всі дзвінки. Окрім цього гарним прикладом є зеркалювання трафіку по аналогії системам IPSMDS. Також RSPAN дає змогу передавати трафік до віддаленого комутатора.

Kismet - це багатофункціональна безкоштовна утиліта для роботи з безпроводовими мережами Wi-Fi. В основному ця програма використовується для спостереження та аналізу мереж IEEE 802.11x. Умовно, завдання, що вирішуються за допомогою Kismet, можна розділити на дві сфери: аналітика та захист. У першому випадку накопичені відомості повинні оброблятися сторонніми програмами, а в другому Kismet працює як детектор різного роду мережевих атак, тобто як аналізатор мережевого трафіку [4-5].

Wireshark - це відомий інструмент для захоплення та аналізу мережевого трафіку, являє собою стандартне програмне забезпечення. Ця програма допомагає мережевим адміністраторам виконувати дослідження мережевих додатків, протоколів, пакетів щоб знайти проблеми у роботі мережі, і що важливо, з'ясувати причини цих проблем. Деякі спеціалісти використовують її для пентесту, аналіз мережі на зловживання мережею, переповнення мережі та інше.

Процес виявлення кібератаки в мережі найчастіше заснований на порівнянні характеристик трафіку на незначному відрізку часу вторгнення з відповідною легітимною трасою трафіку, розглянутої за тривалий період часу. У цьому випадку, якщо виявлено значні «сплески» трафіку, виникає необхідність знайти такий метод виявлення кібератак, у якому похибка виявлення буде мінімальною.

Для протидії зловмисникам сьогодні створено низку складних систем виявлення вторгнень IDS/IPS (Intrusion detection system/Intrusion prevention system), міжмережеві екрани, аналізатори трафіку на базі утиліт та сніферів (tcpdump, Wireshark, Snort), протоколи SNMP (NetFlow). Для оцінки трафіку ці системи використовують різні методи, серед яких: статистичні методи, інтелектуальний аналіз даних, штучний інтелект, вейвлет-аналіз та інші [4-5].

2.6 Обгрунтування необхідності дослідження аномалій мережного трафіку DDoS-атак

Згідно з вищевикладеним слід зазначити, що актуальною є завдання виявлення «сплесків» мережевого трафіку атак типу DDoS (SYN-Flood, ICMP-Flood, UDP-Flood), рішення якої розглянуто у роботах авторів [4-5].

Один із широко використовуваних підходів до виявлення DDoS-атак з використанням марківських моделей запропонований у роботі [6]. Використовуючи аналіз шаблонів атаки за допомогою прихованої марківської моделі, отримано прогноз атак, хоча частина трафіку атаки може бути втрачена. У роботі [7] авторами запропоновано метод, заснований на байєсовських мережах, за допомогою якого було покращено прогноз виявлення вторгнень. Такий похід значно ускладнював визначення атаки і при цьому не вдалося уникнути хибних прогнозів. У роботі [7] на основі байєсівської мережі запропоновано раннє прогнозування з використанням кореляції. Проте така реалізація не є універсальною і не дозволяє прогнозувати всі види атак.

Для виявлення атак з урахуванням розширеного набору ознак у роботі [8] використано машинне навчання ML/DL. Така реалізація, як правило, вимагає навчання та значних часових витрат. Використання вейвлет-перетворень для визначення DDoS-атак у роботі [9] дозволяє значно покращити результати виявлення та прогнозування загроз. Автори роботи [10] при порівняльному аналізі застосовуваних методів виявили, що метод регресійного дерева кращим для виявлення атак у трафіку. Переваги використання штучного інтелекту обґрунтовані авторами у роботі [11], а способи забезпечення мережевої кібербезпеки у роботах [11-12].

2.7 Огляд літератури щодо тематики дослідження та постановка завдання дослідження

Виявлення DDoS-атак є надзвичайно складним та трудомістким завданням. Існуючі системи виявлення вторгнень вимагають постійного оновлення набору правил для завчасного виявлення загроз [12]. З кожним роком пропускна здатність інтернет-каналу зростає, тому застосування лише апаратних засобів детектування мережного трафіку недоцільно. При цьому використання класичних засобів захисту, таких як міжмережеві екрани на сьогоднішній день є малоефективним [7] через використання певного набору правил, відповідно до якого здійснюється фільтрація всіх даних. Тим не менш, навіть періодичного оновлення набору правил може бути недостатньо для того, щоб система завжди залишалася в актуальному стані [4-5]. Найчастіше для детектування таких атак проводиться аналіз аномалій мережевого трафіку, тобто за штатних умов роботи мережі ведеться пошук відхилень від контрольних показників трафіку.

Значна кількість робіт присвячена дослідженню методів захисту від DDoS-атак [7-12]. Загалом багато авторів мають спільну думку, щодо використання статистичних методів та реалізацій на базі апаратно-програмних засобів для детектування DDoS-атак на L2-L4 рівні OSI. Однак, DDoS-атаки на L7 рівні OSI, як правило, мають значну кількість мережних пакетів, як надходять з великою кількістю атрибутів. Тому на цьому рівні необхідно розділяти запити від ботів та реальних користувачів. Для рішення цієї задачі більш доцільним є використання методів машинного навчання щодо виявлення та попередження DDoS-атак [12].

В результаті DDoS-атак функціонування мережевих ресурсів виявляється утрудненим або зовсім неможливим внаслідок надлишкового трафіку. Наприклад, Flood-атака, використовуючи слабкість протоколу DNS, збільшує кількість

трафіку на об'єкт, що атакується [11]. Метою SYN-атак є переповнення стека TCP операційної системи [12], а атаки Low-Rate спрямовані на переповнення буфера маршрутизатора або сервера, що згодом призводить до відмови обробки легітимного трафіку [12].

Використання мережевих екранів для аналізу і фільтрації мережного трафіку ще до попадання на сервер, дозволяє обробити пакетні фільтри, що обробляє пакети по одному. Аналіз проводиться для протоколів TCP, UDP. Вилучені з вхідного пакета дані аналізуються шляхом порівняння із заданим набором правил, залежно від результату порівняння застосовується правило блокування: пакета або перепустка до мережі з відправкою повідомлення джерелу пакета про подію [12]. Пізніше, зі збільшенням каналу передачі даних у комп'ютерній мережі, розробкою та використанням великої кількості веб-ресурсів та додатків, зросла складність та витонченість мережевих атак. З'являються комбіновані атаки, які використовують кілька мережевих протоколів одночасно. Тому є необхідним виконання захисту інформації для всіх відомих атак, передбачати раніше невідомі атаки та підтримувати всі системи безпеки в актуальному стані.

2.8 Вихідні дані трафіку DDoS-атак

Мережевий трафік визначається як односпрямована послідовність сукупності пакетів, що передаються, з деякими загальними властивостями, які проходять через мережевий пристрій. Записи трафіку включають різну інформацію: IP-адреси, кількість пакетів і байтів, часову мітку, тип обслуговування, порти, інтерфейси введення і виведення та іншу інформацію. Дані сеансу, що містять IP-адреси клієнта, номери порту клієнта, IP-адреси сервера, номери порту сервера та протоколу, включеного в дані потоку, важливі для ідентифікації з'єднання. Досліджуючи кореляцію запитів та відповідей клієнт-серверних діалогів, аналізуючи трафік, можна знайти значні траєкторії та зв'язок мережевої аномалії з тією чи іншою подією.

Різрізняють декілька типів DDoS-атак, таких як, IP-fllod, SYN-flood, UDP-flood, TCP-flood, Ping of Death та інші, які характеризуються різким зростанням обсягу трафіка. Частіше атаки відуваються на рівнях L3, L4 та L7 моделі OSI. На рівнях L3 та L4 виконуються атаки мережного (network DDoS) та транспортного (transport DDoS) рівнів. Ефективність протидії DDoS-атакам визначається часом їх визначення та відповідним визначенням певної стратегії керуванням трафіком, а саме, зміною приницпів матршрутизації, коли більша частина підозрілих пакетів

перенаправлюється з сервера, а інші оброблюються частинами [3]. Також, певним рішенням щодо запобігання DDoS-атакам є фільтрація вхідних даних та відхилання пакетів та з'єднань зі шкідливим трафіком.

Для дослідження характеристик трафіку DDoS-атак використаємо відкритий ресурс https://www.kaggle.com [13] мережних атак, https://www.kaggle.com/datasets/jacobvs/ddos-attack-network-logs/data, який містить позначені мережеві журнали різних типів мережевих атак, таких як, UDP-Flood, Smurf, SIDDOS, HTTP-FLOOD та звичайний трафік. Дані містять наступні атрибути:

@attribute SRC_ADD числове
@attribute DES_ADD числове
@attribute PKT_ID числовий
@attribute FROM_NODE числовий
@attribute TO_NODE числовий
@attribute PKT_TYPE {tcp,ack,cbr,ping}
@attribute PKT_SIZE цифровий
@attribute ФЛАГИ {-------,---A---}
@attribute FID числове
@attribute SEQ_NUMBER числове
@attribute NUMBER_OF_PKT числовий
@attribute NUMBER_OF_BYTE цифровий
@attribute NODE_NAME_FROM {Switch1,Router,server1,router,clien-4,client-2,Switch2,client-5,clien-9,clien-2,clien-1,clien-14,clien-5,clien-11, клієнт-13, клієнт-0, перемикач1, клієнт-4, клієнтhttp, клієнт-7, клієнт-19, клієнт-14, клієнт-12, клієнт-8, клієнт-15, список веб-серверів, клієнт-18, клієнт-1, switch2,клієнт-6,клієнт-10,клієнт-7,веб-кеш,клієнт-10,клієнт-15,клієнт-3,клієнт-17,клієнт-16,клієнт-17,клієнт-18,клієнт-12,клієнт- 8,клієнт-0,клієнт-16,клієнт-13,клієнт-11,клієнт-6,клієнт-3,клієнт-9,клієнт-19,http_клієнт}
@attribute NODE_NAME_TO {Маршрутизатор, сервер1, комутатор2, комутатор1, клієнт-1, клієнт-5, клієнт-7, комутатор1, клієнт-11, клієнт-15, клієнт-13, клієнт-3, клієнт-9, клієнт-6, маршрутизатор,клієнт-4,клієнт-14,перемикач2,клієнт-8,клієнтhttp,веб-кеш,клієнт-10,клієнт-12, список веб-серверів,клієнт-0,клієнт-2,http_клієнт,клієнт-13,клієнт-9,клієнт- 1,клієнт-19,клієнт-4,клієнт-17,клієнт-7,клієнт-3,клієнт-12,клієнт-2,клієнт-18,клієнт-16,клієнт-17,клієнт-0,клієнт-16, клієнт-18, клієнт-5, клієнт-11, клієнт-14, клієнт-8, клієнт-6, клієнт-10, клієнт-19, клієнт-15}
@attribute PKT_IN числовий

@attribute PKT_OUT числовий
@attribute PKT_R числовий
@attribute PKT_DELAY_NODE числовий
@attribute PKT_RATE числове значення
@attribute BYTE_RATE числове значення
@attribute PKT_AVG_SIZE цифровий
@attribute ВИКОРИСТАННЯ числове
@attribute PKT_DELAY числове
@attribute PKT_SEND_TIME числове
@attribute PKT_RESEVED_TIME числове
@attribute FIRST_PKT_SENT числовий
@attribute LAST_PKT_RESEVED числовий
@attribute PKT_CLASS {Normal,UDP-Flood,Smurf,SIDDOS,HTTP-FLOOD}

Трафік DDoS-атак визначається як від об'єму передаваємих даних кожного з пристроїв, так і загальною їх кількістю. Природно, що при різних способах класифікації оцінки величини цього трафіку можуть істотно відрізнятися.

Під трафіком DDoS-атак будемо розуміти характеристики потоку пакетів, які вироблені серверами мережі.

Потоки пакетів DDoS-атак будемо описувати параметрами трафіку (інтенсивністю надходження пакетів, розмір пакетів) так і характеристиками потоку, такими як розподіл інтервалів часу між пакетами.

Потоки будемо вважати випадковими та в деяких послугах – детермінованими. У більшості випадків, а також в даній роботі, трафік в мережі буде характеризуватися швидкісними параметрами (пропускною здатністю), часовими параметрами (часом затримки доставки пакетів) і ймовірносними параметрами (ймовірністю втрат пакетів).

Розглянемо характеристики інтенсивності лігітімного трафіку мережі та DDoS-трафік отриманий за допомогою Dataset/ddos-attack-network [13], який показаний на рис. 2.1.

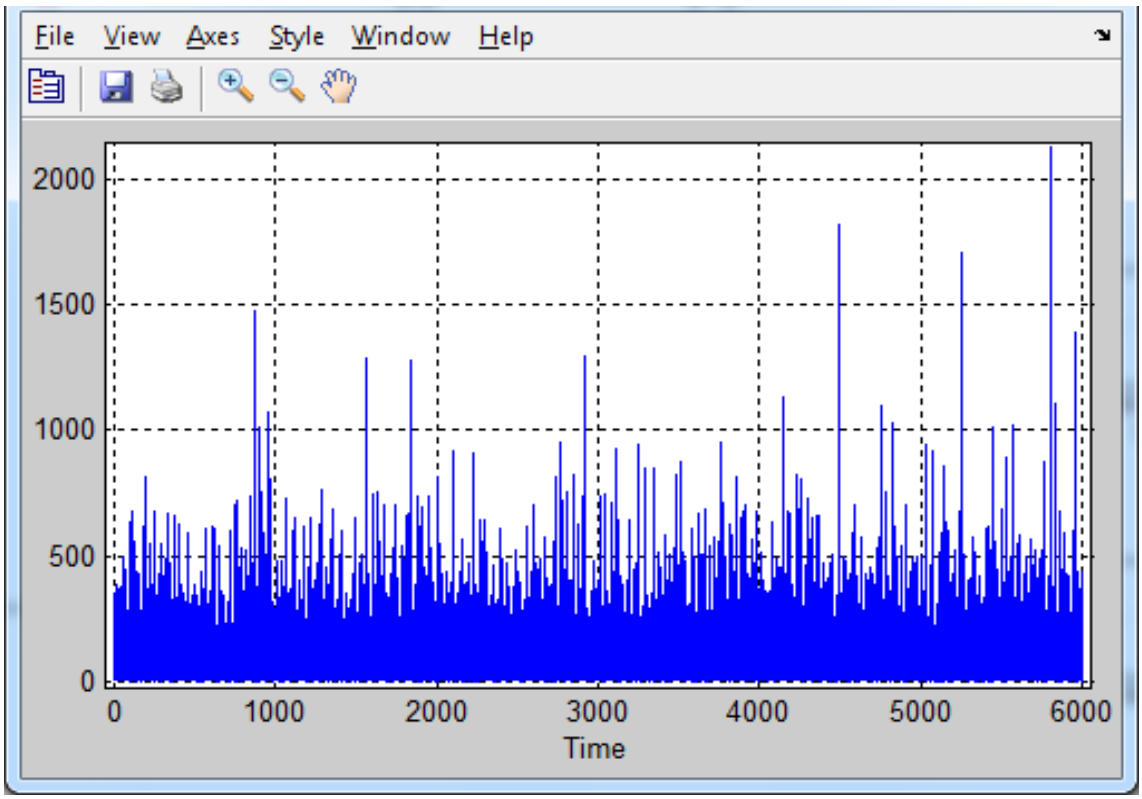


Рисунок 2.1 – Трафік, отриманий за допомогою Dataset/ddos-attack-network

Отримані дані вихідного трафіку використовували агреговані п’ятихвилинні інтервали, середні значення наступних величин: кількість байт в секунду, кількість пакетів в секунду, кількість потоків в секунду, величина середнього розміру TCP-пакета. Наявність різких «сплесків» відповідає певним групам аномалій.

Огляд вихідного трафіку на рис. 2.1 дозволяє зробити наступні висновки:

– трафік характеризується значною кількістю «сплесків» інтенсивності пакетів,

– трафік має ознаки самоподібності зі значенням параметра Херста H = 0,8.

Після збору і аналізу трафіку можна надати наступні типи аномалій в мережі:

– обладнання: відмова обладнання або тимчасова невірна настройка обладнання, відключення;

– DDoS-атаки, зокрема типу flood, такі як SYN-flood, ICMP-flood,

– перевантаження на мережі, збільшення інтенсивності вихідного трафіку Ftp–сервера внаслідок появи на ньому популярного контенту;

– інші аномалії, які не належать ні до проблем на мережі, ні до атак і перевантажень.

Вищезазаначений огляд дозволяє зробити висновки про те, що трафік в

мережі є пакетним і різнорідним, так як формується безліччю різних за своїми характеристиками джерел послуг та мережевих додатків і характеризується проявом самоподібних властивостей, що мають на увазі наявність довгострокової залежності між моментами надходження пакетів, визначається функцією кореляції в різні моменти часу.

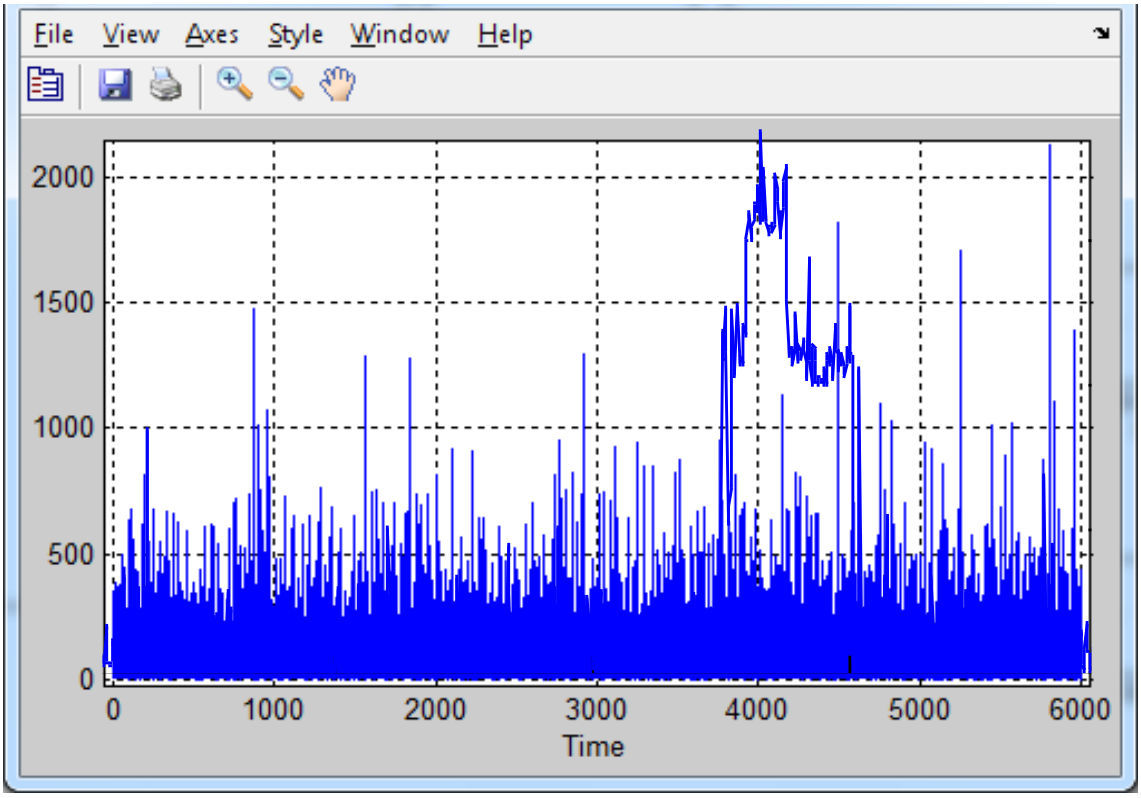


Рисунок 2.2 – Трафік DDoS-атаки виду ICMP-flood, отриманий за допомогою Dataset/ddos-attack-network

Неважко бачити на рис. 2.2, що трафік має самоподібні властивості та подібні часті «сплески» трафіку, які властиві для відеотрафіку. Частіше такий трафік має високе значення параметру Херсту H = 0,9. Встановлено, що такий трафік описується законом розподілу Вейбулу. Тоді можливо отримати наступні параметри кривої розподілу Вейбулу, а саме, де α – параметр форми кривої, $0<\alpha<1$, який визначається виразом

$$\alpha = 2 - 2H = 2 - 2\times 0{,}9 = 0{,}2.$$

H – параметр Херста, H = 0,9, $0{,}5 \le H \le 1$,

Параметр розподілу Вейбула β, $\beta>0$, визначається як

$$\beta = \left[\lambda \Gamma\left(1 + \frac{1}{\alpha}\right) \right]^{\alpha}, \beta = 0{,}00005$$

Розгянемо трафік, який показано на рис. 2.1 з DDoS-атакою виду UDP Flood, ця DoS-атака спрямована на вичерпання обчислювальної потужності мережі та досягнення граничних значень смуги пропускання. UDP-потоки зазвичай мають дуже високу інтенсивність і фактично складаються із запитів, що викликають неприпустимі відповіді (UDP типу 0). UDP -потоки, якщо вони не заблоковані засобами захисту від DDoS -атак, можуть порушити безпеку внутрішньої мережі. Вони також можуть генерувати трафік, що складається з багатьох відповідей на UDP -запити.

Розглянемо використання статистичного методу дослідження аномалій для визначення трафіку DDoS-атаки виду UDP -flood (рис 2.2) [15].

2.9 Обробка статистичних даних DDoS-атаки виду UDP-flood

Для проведення досліджень характеристик аномального трафіку DDoS-атак на базі отриманих статистичних даних необхідно визначити, який теоретичний закон розподілу трафіку узгоджується зі статистичним розподілом, отриманим для реального трафіку DDoS-атаки.

Для оцінки результатів трафіку DDoS-атаки використовують побудову гістограм за результатами отриманих статистичних даних вимірювання трафіку.

Для побудови графіку використаємо дані трафіку, які показані на рис. 2.1. Використовуючи Dataset/ddos-attack-network, отримаємо значення кількісті пакетів трафіку, які отримані на за визначений термін часу [15].

Для побудови гістограми по осі X використовуємо значення «Час дослідження трафіку DDoS-атаки», дану вісь розіб’ємо на 50 значень відповідно до отриманих статистичних даних на рис. 2.1, а по осі У використовуємо дані про кількість ICMP-пакетів, які надійшли. Для побудови даної осі, спочатку було визначене максимальне значення кількості пакетів, а потім вісь була розбита від 0 до максимального значення, з кроком 20.

Згідно даних, які отримані на рис. 2.1 побудовано відповідну гістограму, яка відображає інформацію щодо кількісті UDP-пакетів, які надходять.

Основними завданнями обробки результатів вимірювання є: обчислення оцінки вимірюваного параметра та обчислення ступеня достовірності цієї оцінки. При статистичних вимірах за найбільш ймовірне значення вимірюваної величини

приймається середнє арифметичне значення. Ступінь розбіжностей між собою окремих значень досліджуваного ознаки характеризується середньоквадратичним відхиленням [15].

Нехай над випадковою величиною X проводиться n незалежних дослідів, і нехай $X_к$ – випадкова величина, яка дорівнює значенню, що заздалегідь невідомо, прийнятому величиною X у $к$-ому досліді. Тоді сукупність незалежних випадкових величин [15]:

$$X_1, X_2, \ldots X_n \tag{2.1}$$

що мають той самий закон розподілу, що і для X, можна розглядати як випадкову вибірку.

Нехай a – деяка числова характеристика величини X, що нам невідома. Тоді оцінка цієї величини ã, що обчислена на основі сукупності величин (2.1), буде являти собою деяку функцію від $X_1, X_2, \ldots X_n$, і отже саме є випадковою величиною. Закон розподілу $\tilde{a}$ залежить від обраного виду функції f, від закону X і від обсягу вибірки n.

З закону великих чисел відомо, що середнє арифметичне спостереження значень (2.1) випадкової величини X при необмеженому збільшенні числа іспитів n збігається за ймовірністю до математичного спостереження m_x цієї величини. Тому природно за оцінку m_x взяти функцію [15]:

$$\tilde{m}_x = \frac{1}{n}\sum_{i=1}^{n} X_i \,, \tag{2.2}$$

Розглянемо тепер дисперсію величини X [15]:

$$D_x = M\left[(x - m_x)^2\right], \tag{2.3}$$

Якщо в цьому виразі математичне сподівання замінити його оцінкою, то ми отримаємо оцінку дисперсії [15]:

$$\overline{D}_x = \frac{1}{n}\sum_{i=1}^{n} (x_i - m_x)^2 \,, \tag{2.4}$$

В даному випадку мали місце вибіркові спостереження за навантаженням у

вибраний день в трьох різних напрямках. Для отримання у результаті вибіркових статистичних спостережень різноманітних числових значень визначені середні значення та визначена степінь розбіжності окремих значень поміж собою та поміж окремих значень та середнім значенням.

Дана розбіжність характеризується середньоквадратичним відхиленням σ, по розміру якого робиться висновок про розкид окремих значень навколо їх середнього значення. Через степінь розкиду розрахована точність спостережень, які відбулися.

Для оцінки результатів вимірювання використаємо розрахунки абсолютної та граничної похибки. Значення середньої абсолютної похибки μ_x репрезентативності випадкової повторної вибірки наближено розраховується за формулою [15]:

$$\mu_x \approx \frac{\sigma(x)}{\sqrt{n}} \tag{2.5}$$

де $\sigma(x)$ – середньоквадратичне відхилення вибіркової сукупності, n – обсяг вибірки.

Значення $\sigma(x)$ було отримано із дослідження числових характеристик випадкової величини X: математичного очікування m_x, дисперсії D_x, та відповідні їм статистичні аналогії, що будуються на основі конкретних результатів дослідів величини X.

Кожну із статистичних характеристик слід припустити рівною тому значенню відповідної оцінки параметра $\tilde{a}$, що ця оцінка приймає для даної вибірки, тобто припустити [15]:

$$\tilde{a} = f(x^{(1)}, x^{(2)}, ..., x^{(n)}). \tag{2.6}$$

Таким чином, математичне очікування, дисперсія та середньоквадратичне відхилення m_x, D_x, $\sigma(x)$ визначаються такими виразами [15]:

$$m_x = \frac{1}{n}\sum_{i=1}^{n} x^i, \tag{2.7}$$

де x^i – поточне значення параметра, n – обсяг вибірки.

$$D_x = \frac{1}{n-1}\sum_{i=1}^{n}(x^i - m_x)^2 , \qquad (2.8)$$

де n – обсяг вибірки, m_x – значення математичного очікування.

$$\tilde{\sigma}_x = \sqrt{\tilde{D}_x} . \qquad (2.9)$$

де D_x – значення дисперсії.

Вихідні дані для побудови гістограми трафіка DDoS-атаки занесено до табл. 2.10. Математичне очікування та дисперсію та інші параметри гістограми трафіка DDoS-атаки занесемо до табл. 2.11.

Оскільки склад вибіркової сукупності є випадковим, то середнє значення вибірки в окремих вибірках може значно відрізнятися від дійсного значення даної величини.

Враховуючи, що вибірки розподілені по визначеному закону, з певною ймовірністю можна стверджувати, що відхилення вибіркової середньої від дійсного середнього значення величини не перевищить заданої величини Δ_x, яка називається граничною помилкою вибірки, а ймовірність - довірчою ймовірністю [15]. Граничну помилку вибірки Δ_x визначимо із співвідношення [15]:

$$\Delta_x = z\mu_x = \frac{z\sigma}{\sqrt{n}}, \qquad (2.10)$$

де z – коефіцієнт, який залежить від ймовірності, з якої можливо гарантувати певне значення граничної помилки, σ – середньоквадратичне відхилення вибіркової сукупності, n – обсяг вибірки.

Згідно [15], оберемо значення ймовірності, яке становить $p(z)$=0,99 та коефіцієнт, який залежить від ймовірності, з якого гарантується гранична похибка z=2,6. Зі збільшенням коефіцієнту z ймовірність $p(z)$ наближається до одиниці. Гранична помилка вибірки Δ_x показує, що з певної ймовірністю відхилення вибірки не перевищують заданої величини.

У відповідності зі знайденою граничною помилкою статистичних даних про навантаження, кількості викликів та тривалості зайнятості при ймовірності гарантованого значення граничної помилки P(z)=0,99. Для визначення граничної помилки навантаження та числа викликів обсяг вибірки складає 50 спостережень.

Таблиця 2.10 – Розраховані параметри для побудови гістограми трафіку DDoS-атаки

i	1	2	3	4	5	6	7	8	9	10
Інтервали $(x_i; x_{i+1})$	⋮	100-120	120-140	140-160	⋮	700-740	740-760	760-780	⋮	980-1000
Кількість інтервалів m_i	-	6	7	7	-	2	2	1	-	1
Відносна частота p_i	-	0,051	0,050	0,042	-	0,002	0,001	0,003	-	0,000
x_i	-	110,0	130,0	150,0	-	710,0	730,0	770,0	-	990,0
x_i^2	-	12100,0	16900,0	22500,0	-	504200,0	532800,0	592800,0	-	980100,0
$x_i^2 p_i$	-	621,67	847,86	946,11	-	826,62	737,20	165,08	-	0,0

Таблиця 2.11 – Додаткові параметри, необхідні для дослідження статистичних даних трафіку DDoS-атаки

$\tilde{m}_x$	$\tilde{D}_x$	σ	N	μ	Δx	β
159,13	16895,10	119,93	50	0,19	7,7	12,53%

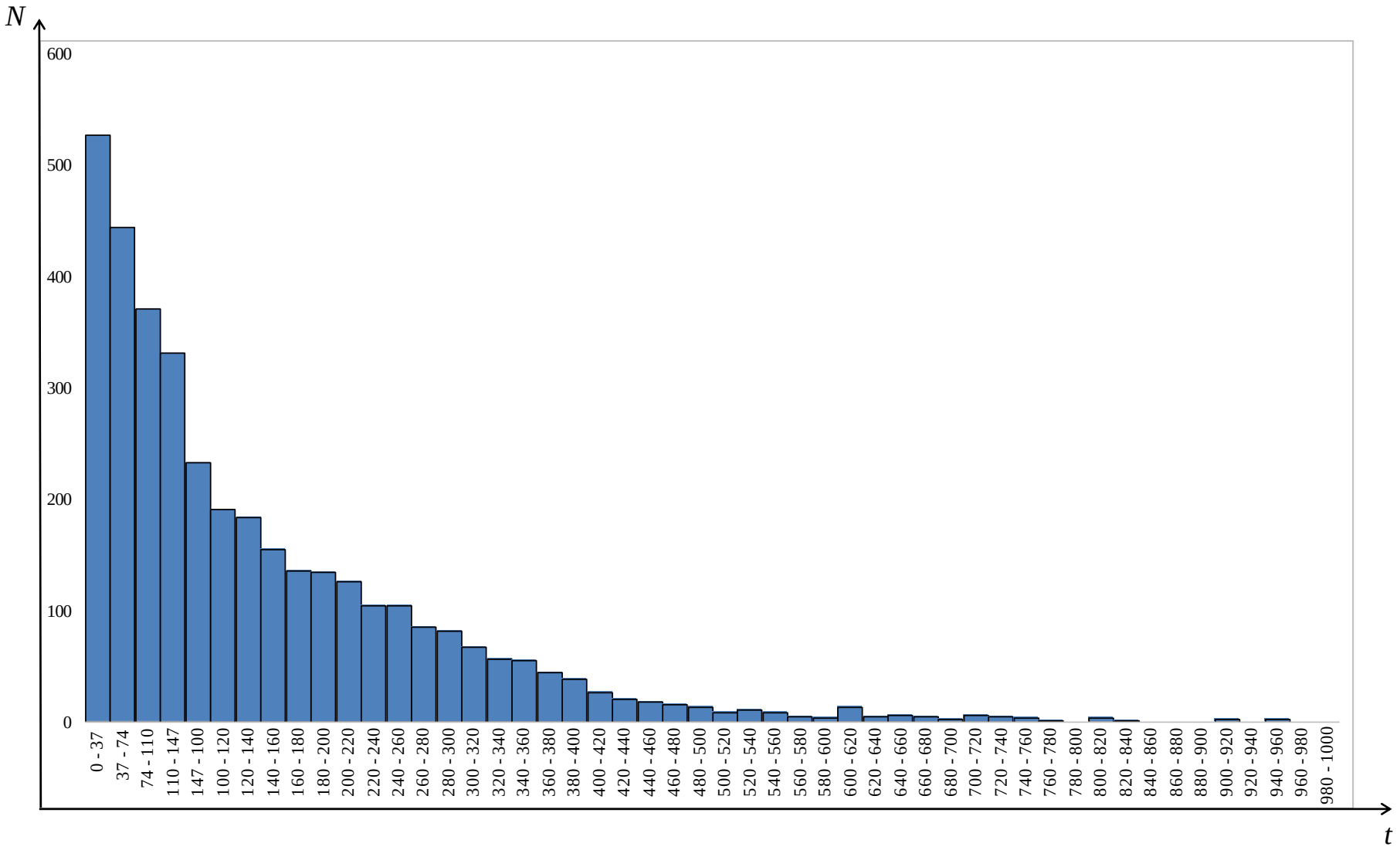


Рисунок 2.4 – Гистограма трафіку DDoS-атаки

Розрахунок граничної помилки виміру показали, що відносна похибка, знаходиться із співвідношення [15]:

$$\beta = \frac{\Delta x}{M_x}, \tag{2.11}$$

де Δ_x – гранична похибка вибірки, Mx – математичне очікування.

Зробимо розрахунок даної величини для отриманих даних та занесемо до табл. 2.10. Отримані значення відносної помилки знаходяться в межах 11-14%, отже отримані дані можуть вважатися достатньо вірними з ймовірністю p(z)=0,99.

Крім помилок в визначені достовірності кінцевого результату залежить від точності розрахунку при обробці результатів виміру. Згідно правил похибка виміру повинна бути приблизно на порядок (тобто в 10-15 разів) нижче сумарної помилки виміру [15].

2.10 Вибір закону розподілу апроксимації статистичних даних трафіку DDoS-атак

При обробці дослідницьких даних трафіку DDoS, отриманих у результаті спостережень величини X в серії дослідів, часто доводиться вирішувати питання про те, як на основі цих даних підібрати теоретичну криву розподілу, яку можна було б прийняти як передбачуваний закон розподілу величини X. Така задача називається задачею вирівнювання (згладжування) статистичних розподілів [15].

Узагальненому випадку для розв’язування цієї задачі, що полягає в аналітичному поданні емпіричних розподілів, користуються методом найменших квадратів, відповідно до якого вважається, що найкращим наближенням до емпіричної залежності у даному класі функції є таке, при якому сума квадратів відхилень перетворюється на мінімум. Також, при використанні статистичних даних важливо визначити закон розподілу. Це необхідно зробити для оцінки отриманих даних, при цьому потрібно враховувати наступний фактор, такий як вибірка [15].

Вибірка – це множина об'єктів, подій, зразків або сукупність вимірів, за допомогою визначеної процедури вибраних з статистичної популяції або генеральної сукупності для участі в дослідженні. Зазвичай, розміри популяції дуже великі, що робить прийняття до уваги всіх членів популяції непрактичним або неможливим. Вибірка являє собою множину або сукупність певного обсягу,

члени якої збираються і статистичні характеристики обчислюються таким чином, що в результаті можна зробити висновки або екстраполяцію із вибірки на всю популяцію або генеральну сукупність [15].

Частіш за все статистичні дані дослідження трафіку DDoS виконують за допомогою закону розподілу. Розглянемо можливість використання певного закону розподілу для апроксимації статистичних даних, які описані гістограмою на рис. 2.4.

Слід зауважити, що при побудові гістограми DDoS трафіку на заданому проміжку, збільшуючи кількість розбиття цього проміжку, збільшується тривалість часткових інтервалів, яка буде прямувати до нуля. Замінюючи ступінчасту лінію отриманої гістограми безперервної кривої, одержимо графік щільності розподілу випадкової величини [15].

Використовуючи дані розподілу вибірки, побудуємо емпіричну функцію розподілу характеристики надходження пакетів трафіку DDoS-атаки в проміжку [1;1000]. По виду цієї кривої робимо припущення про те, що розподіл виконується за законом розподілу Вейбулла.

Після того, як висловлено припущення про теоретичному законі розподілу трафіку, виникає питання про те, якою мірою воно узгоджується зі статистичним розподілом, отриманим на основі моделювання трафіку. Використовуючи критерій Пірсона $\chi 2$ [15], неважко перевірити узгодженість між теоретичним і статистичним розподілами.

Розподіл Вейбулу може бути заданий диференціальною функцією розподілу вигляду [15]:

$$f(x)=\begin{cases}\alpha\beta\, x^{\alpha-1}e^{-\beta x^{\alpha}}, & x\geq 0\\ 0, & x\leq 0\end{cases}, \qquad (2.12)$$

де α – параметр форми кривої розподілу Вейбулу, $0<\alpha<1$, який визначається, відповідно [6], виразом

$$\alpha = 2 - 2H,$$

H – параметр Херста, $0{,}5 \leq H \leq 1$,

$\beta = \left[\lambda\Gamma\left(1+\frac{1}{\alpha}\right)\right]^{\alpha}$ – параметр розподілу Вейбула, $\beta>0$,

Γ – гама-функція Ейлера вигляду $\Gamma(k) = \int_0^{+\infty} t^{k-1} e^{-1} dt$,

λ – інтенсивність надходження пакетів на обслуговування.

Інтегральна функція розподілу Вейбула має вигляд [6]:

$$F(x) = 1 - e^{-\beta x^{\alpha}}. \quad (2.13)$$

Математичне очікування розподілу Вейбулу має вигляд [6]:

$$M = \beta^{-\frac{1}{\alpha}} \Gamma\left(1 + \frac{1}{\alpha}\right). \quad (2.14)$$

Дисперсія розподілу Вейбула [6]:

$$D = \beta^{-\frac{2}{\alpha}} \left[\left(\Gamma\left(1 + \frac{2}{\alpha}\right) - \Gamma^2\left(1 + \frac{1}{\alpha}\right)\right)\right], \quad (2.15)$$

де $\Gamma(x)$ – гама-функція Ейлера.

Досліджено трафік при нормальному функціонуванні мережі та при впливі DDoS-атаки типу UDP-flood. У цьому розглядалася послідовність інтервалів між надходженням пакетів. Були визначені функції густини розподілу ймовірності, кореляційні функції та коефіцієнти достовірності для обох випадків. При дії DDoS-атаки типу UDP-flood статистичні властивості послідовності суттєво змінюються, тобто відхилення статистичних властивостей може бути ознакою DDoS-атаки.

Отримані значення кореляційних функцій передбачається використовувати під час аналізу даних із прийняття рішення про наявність атаки (аномального трафіку). Крім того, результати дослідження є основою подальшого аналізу підозрілого трафіку. При цьому може здійснюватися моделювання різних типів DDoS-атак.

3 ВИКОРИСТАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ДЕТЕКТУВАННЯ АНОМАЛІЙ МЕРЕЖНОГО ТРАФІКУ DDoS-АТАК

3.1 Обгрунтування необхідності дослідження аномалій мережного трафіку DDoS-атак за допомогою методів машинного навчання

Постійне збільшення використання мережевої зв'язку в останні роки призвело до збільшення ризику компрометації інформації. Методи вторгнень розвиваються і стають більш витонченими. Отже, класичні системи виявлення вторгнень показують зниження продуктивності при виявленні нових атак [1-2] Системи виявлення вторгнень класифікуються за методиками аналізу на сигнатурні, статистичні (аномальні) і гібридні.

Статистичні методи створюють статистичну модель, що описує нормальний мережевий трафік, і ідентифікують будь-яку ненормальну поведінку, що відхиляється від моделі. Основною проблемою аномальних методів є складність в налаштуванні і велика кількість хибнопозитивних тривог в разі некоректно заданих правил [1-2]. У даній роботі основна увага приділяється системам виявлення вторгнень на основі аномалій, зокрема, щодо виявлення мережевих DDoS-атак.

Статистичні методи виявлення аномалій створюють статистичну модель, що описує нормальний мережевий трафік та поведінку мережі, і ідентифікують будь-яку ненормальну поведінку, що відхиляється від моделі. На відміну від сигнатурних методів, методи на основі аномалій мають ту перевагу, що вони можуть виявляти атаки з нульовим днем, так як нові атаки можуть бути виявлені, як тільки вони відбудуться.

Основною проблемою методів на основі аномалій є складність в налаштуванні і велика кількість хибнопозитивних тривог в разі некоректно заданих правил [1-2]. У даній роботі основна увага приділяється системам на основі статистичного виявлення аномалій, зокрема, щодо виявлення DDoS-атак.

У зв’язку з цим запропоновано підхід до виявлення та захисту від DDoS атак (таких як, UDP-flood, потоків TCP SYN, Ping of Death атак та HTTP flood). Запропонований алгоритм зосереджується на трьох важливих частинах, а саме виявленні, захисті і повідомленні про напади. Запропонований алгоритм виявлення буде перевіряти вхідний трафік, будь то трафік DDoS або звичайний трафік. Якщо вхідний трафік є трафіком DDoS-атак, то запропонований підхід до виявлення буде визначати типи DDoS-атак, такі як UDPflood, потоків TCP SYN, Ping of Death та HTTP flood, засновані на поведінці атаки [3].

Найбільш розповсюджений спосіб організації DDoS-атак – це використання для відправки «хибних» запитів так названих ботнетів, які складаються зі зламаних серверів та комп'ютерів, які мають доступ до мережі Інтернет [3]. Різрізняють декілька типів DDoS-атак, таких як, IP-fllod, SYN-flood, UDP-flood, TCP-flood, Ping of Death та інші, які характеризуються різким зростанням обсягу трафіка. Частіше атаки відуваються на рівнях L3, L4 та L7 моделі OSI. На рівнях L3 та L4 виконуються атаки мережного (network DDoS) та транспортного (transport DDoS) рівнів. Ефективність протидії DDoS-атакам визначається часом їх визначення та відповідним визначенням певної стратегії керуванням трафіком, а саме, зміною приницпів матршрутизації, коли більша частина підозрілих пакетів перенаправлюється з сервера, а інші оброблюються частинами [3]. Також, певним рішенням щодо запобігання DDoS-атакам є фільтрація вхідних даних та відхилання пакетів та з'єднань зі шкідливим трафіком.

3.2 Вихідні дані для дослідження

Вихідними даними для дослідження є вибірки мережевого трафіку, зібрані за допомогою бібліотек [13], які забезпечують взаємодію з драйвером мережного інтерфейсу. У ході виконання роботи було виявлено, що атрибути мережного трафіку при проведенні одного і того ж типу атаки можуть відрізнятися залежно від програмних інструментів, використовуваних зловмисником. Внаслідок цього у процесі збору даних для навчання було задіяно кілька програмних засобів для кожного типу атак (таблиця 3.1), що підвищило різноманітність мережевого трафіку та дозволило точніше виділити інформативні ознаки.

Для дослідження характеристик трафіку DDoS-атак використаємо відкритий ресурс https://www.kaggle.com [13] мережних атак, https://www.kaggle.com/datasets/jacobvs/ddos-attack-network-logs/data, який містить позначені мережеві журнали різних типів мережевих атак, таких як, UDP-Flood, Smurf, SIDDOS, HTTP-FLOOD та звичайний трафік.

Дані datasets/ddos-attack-network/ показані на рис. 3.1. Серед даних трафіку є три DDoS-атаки UDP-flood, дві Smurf та двадцять три атрибути нормального трафіку без атак. Мережний трафік datasets отриманий з вузлів мережі, Routers, Switch та Cliens та розглянутий у програмному середовищі Weka 3.7.1 [14]. Серед характеристик трафіку в datasets відмічено швидкість передавання даних, як в пакетах, так і біт/с, час затримки пакетів, час очікування та час надсилання.

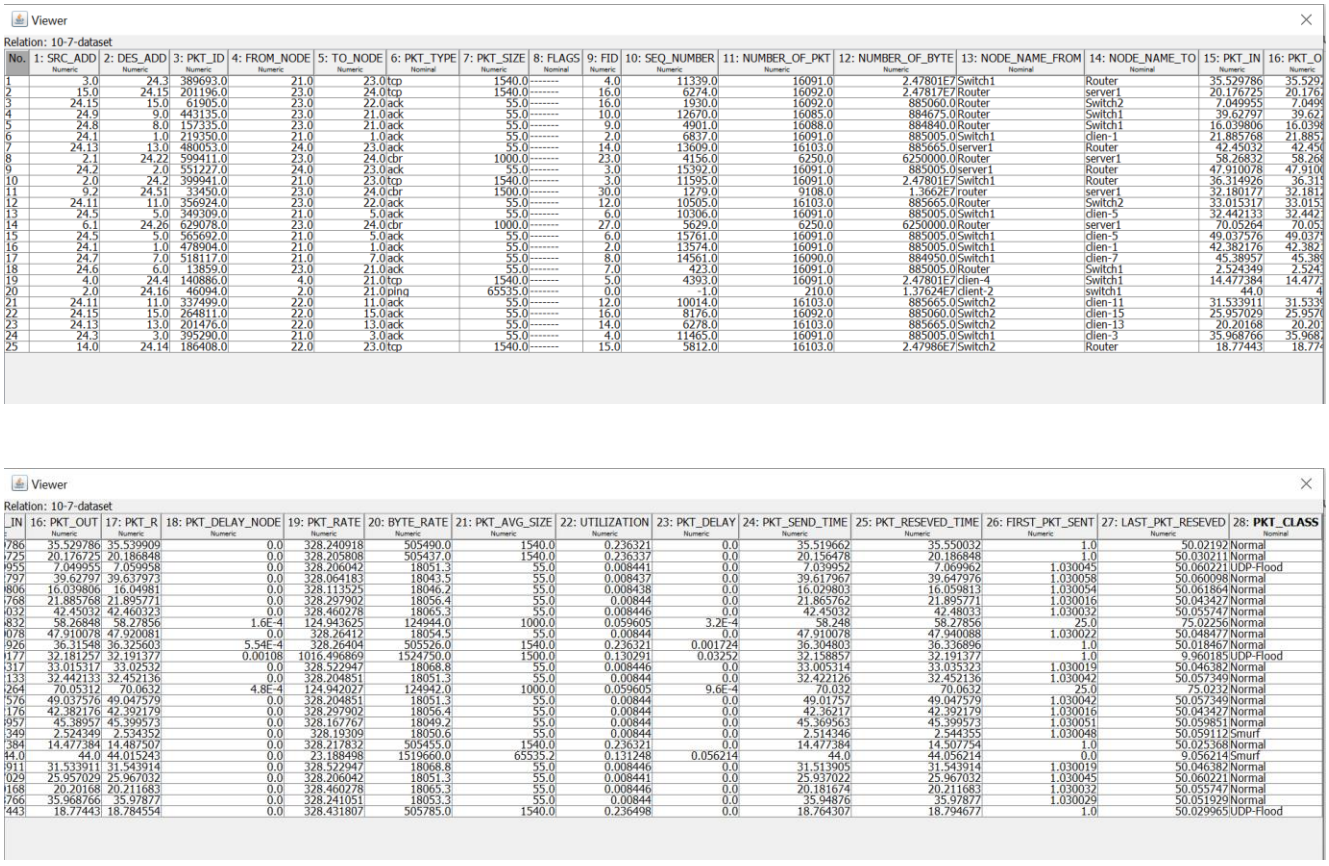


Рисунок 3.1 – Атрибути datasets/ddos-attack-network/

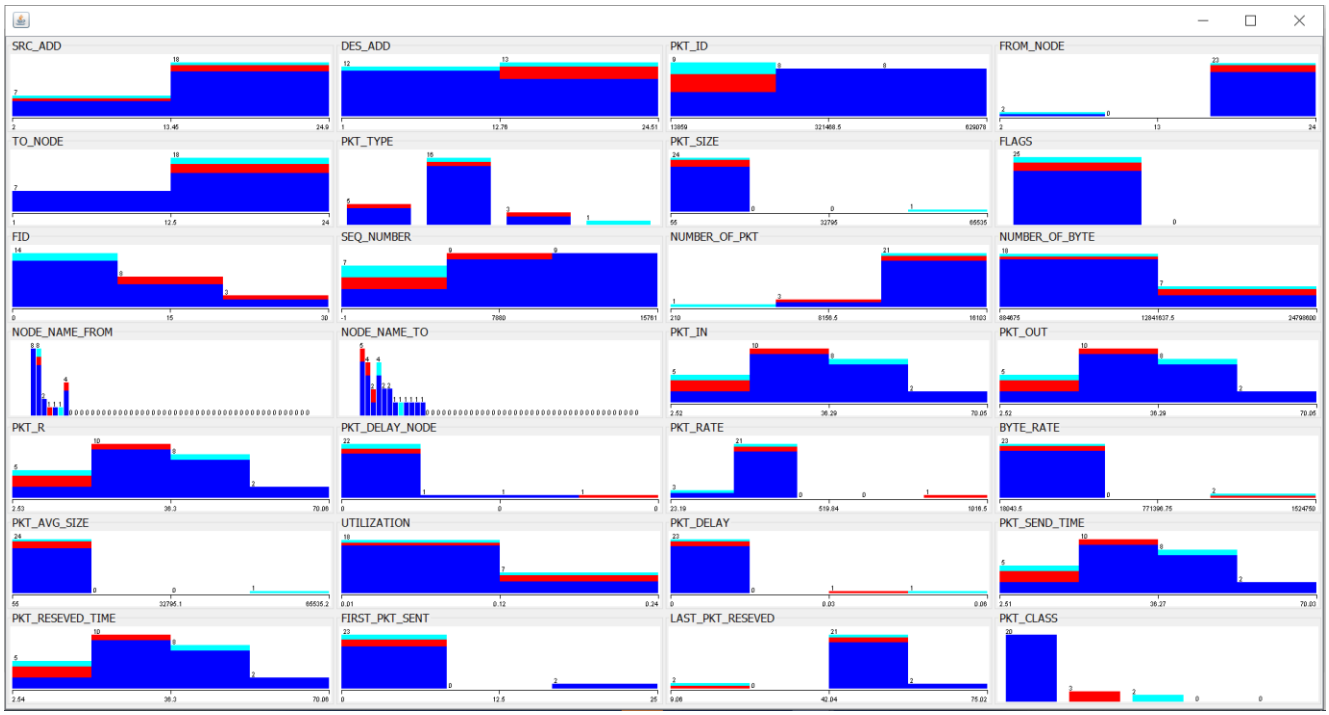


Рисунок 3.2 – Візуалізація datasets/ddos-attack-network/

3.3 Вирішення завдань класифікації

Рішення задачі класифікації передбачає необхідність обрати навчальну вибірку мережного трафіку та знайти функцію $f(x)$ від вихідних параметрів, яка при найменьшій кількості помилково кваліфіцірованих пакетів зможе надати рішення щодо результуючого класу атаки y для будь-яких значень x.

Вихідні дані для вирішення використовують дискретну множину $\{y_1, y_2, \ldots, y_m\}$, яка включає зумовлені значення класів [17]:

$$X \to Y,$$

де где X – це набір векторів атрибутів мережних пакетів, Y – набір видів DDoS-атак. Значення цільової залежності f відомі тільки на об'єктах кінцевої навчальної вибірки [17]:

$$x^m = \{(x_1, x_2), \ldots, (x_m, y_m)\}.$$

При рішенні задачі класифікації необхідно найти такий алгоритм, який

$$a: X \to Y,$$

який здатний класифікувати довільний об'єкт. Як атрибути класифікації повинні використовуватися атрибути пакетів мережевого, транспортного та прикладного рівнів моделі OSI.

Розглянемо рішення задачі класифікації машинного навчання аномалій трафіку DDoS-атак за допомогою наступних алгоритмів [17]:

– алгоритм логістичної регресії,

– алгоритм k-найближчих сусідів,

– алгоритм випадкового лісу.

3.4 Вирішення завдань класифікації методом k-найближчих сусідів

Побудова моделі класифікації на основі алгоритму k найближчих сусідів полягає у запам'ятовуванні навчальної вибірки даних. Для того, щоб зробити прогноз для нового екземпляра даних алгоритм виконує пошук найближчої до неї

точки навчальної вибірки, цим знаходячи «найближчих сусідів». Новому екземпляру надається мітка, що належить найближчої точки навчального набору. Алгоритм дозволляє розглядати не одного найближчого сусіда, а їхня довільна кількість становить k, звідси походить назва алгоритму.

Нехай задано дана навчальна вибірка з парами виду «об'єкт-відповідь» [17]:

$$x^m - \{(x_1, y_2), \dots, (x_m, y_m)\}. \tag{3.1}$$

Нехай на безлічі об'єктів задана функція відстані $p(x, x')$, яка має бути достатньо адекватною моделлю подібності об'єктів, то є чим більше значення цієї функції, тим менш схожими є об'єкти x, x'. Для довільного u розташуємо об'єкти навчальної вибірки x_i в порядку зростання відстані до u [17]:

$$x^m - \{(x_1, y_2), \dots, (x_m, y_m)\}. \tag{3.2}$$

де xi,m - це об'єкт навчальної вибірки, який є i-м сусідом об'єкта u. Аналогічно введемо позначення для yi,u для відповіді на i-му сусіді.

Алгоритм найближчих сусідів у найбільш загальному вигляді виглядає так:

$$a(u) = argmax \sum_{i=1}^{m} [x_{i,m} = y] w(i, u), \tag{3.3}$$

де $w(i, u)$ - це задана вагова функція, що оцінює ступінь важливості i-го сусіда за класифікації об'єкта u [16]. Ця функція має бути невід'ємною і не зростати за i-м.

Проведена класифікація алгоритмом (рис. 3.3) за допомогою пакету Weka 3.7.1 [14], для загальної кількості випадків 25, наведених на рис. 3.1, серед яких, дані трафіку трьох DDoS-атак виду UDP-flood, двох DDoS-атак Smurf та двадцяти атрибутів нормального трафіку без атак [17].

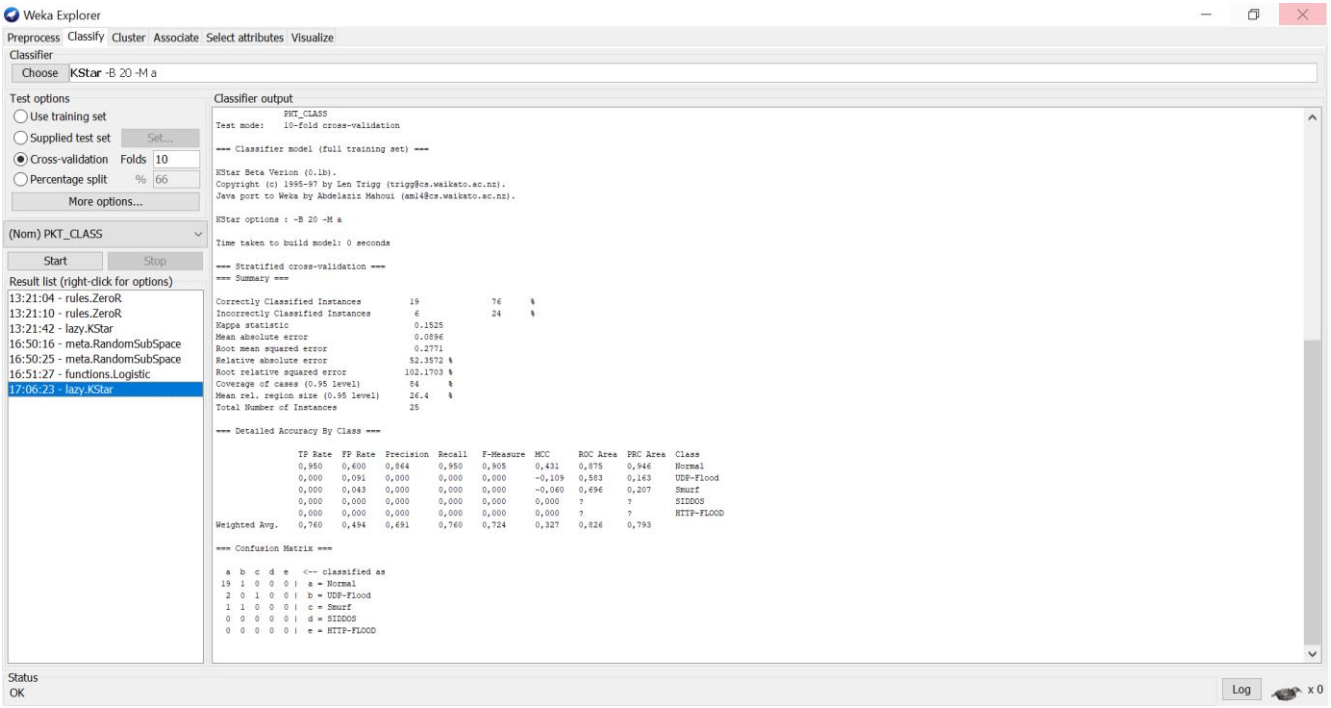


Рисунок 3.3 – Результати класифікації на основі алгоритму k найближчих сусідів

Отримані результати класифікації, надають інформацію про те, що правильно класифіковані всі вибірки мережного трафіку (25 випадків). Отримані результати класифікації наведені в табл. 3.1.

Таблиця 3.1 – Результати класифікації на основі алгоритму k-найближчих сусідів

DDoS-атака	Normal	UDP-flood	Smurf	SI DDoS	HTTP-flood
Normal	19	1	0	0	0
UDP-flood	2	0	1	0	0
Smurf	1	1	0	0	0
SI DDoS	0	0	0	0	0
HTTP-flood	0	0	0	0	0

Класифікація на основі алгоритму k-найближчих сусідів полягає у запам’ятовуванні навчальної вибірки даних. Для аналізу даних трафіку алгоритм виконує пошук найближчої до неї точки навчальної вибірки, тим самим знаходячи k-найближчих [16]. Згідно результатів класифікації наведених в табл. 3.1 можна стверджувати, що дана класифікація трафіку DDoS-атак має похибки.

Алгоритм k-найближчих сусідів має кілька переваг, у тому числі легкість інтерпретації моделі, задовільна якість передбачення, яке може бути одержано без використання великого кількості налаштувань. Більш того, зазвичай цей алгоритм

дозволяє побудувати модель класифікації дуже швидко. Однак за наявності великої навчальної вибірки, моделі потрібен додатковий час, щоб навчитися.

3.5 Вирішення завдання класифікації алгоритмом логістичної регресії

Логістична регресія використовується як метод для двійкової класифікації. Логістична регресія моделює ймовірність подій класифікації з двома можливими результатами і може використовуватися для ідентифікації мережного трафіку DDoS-атак як шкідливого (true) чи ні (false). Основна ідея логістичної регресії полягає в тому, що простір вихідних значень може бути розділений лінійним кордоном (тобто прямою) на дві області, що відповідають класам. Під лінійним кордоном мається на увазі пряма лінія – у разі двох вимірів, у разі трьох вимірів – площина, тощо [16].

Нехай є деяка випадкова величина Y, що може набувати лише двох значень, 0 та 1. Нехай ця величина залежить від деякої множини пояснювальних змінних $x = (1, x_1, \ldots, x_n)^T$. Залежність Y від x_1,…., x_n можна визначити ввівши додаткову змінну y^*, де

$$y^* = \theta^T x = \theta_0 + \theta_1 x_1 + \ldots + \theta_n x_n + \varepsilon.$$

Тоді

$$Y = \begin{cases} 0, & y^* \leqslant 0 \\ 1, & y^* > 0 \end{cases}$$

Проведена класифікація алгоритмом (рис. 3.4) за допомогою пакету Weka 3.7.1 [14], для загальної кількості випадків 25, наведених на рис. 3.4, серед яких, дані трафіку трьох DDoS-атак виду UDP-flood, двох DDoS-атак Smurf та двадцяти атрибутів нормального трафіку без атак.

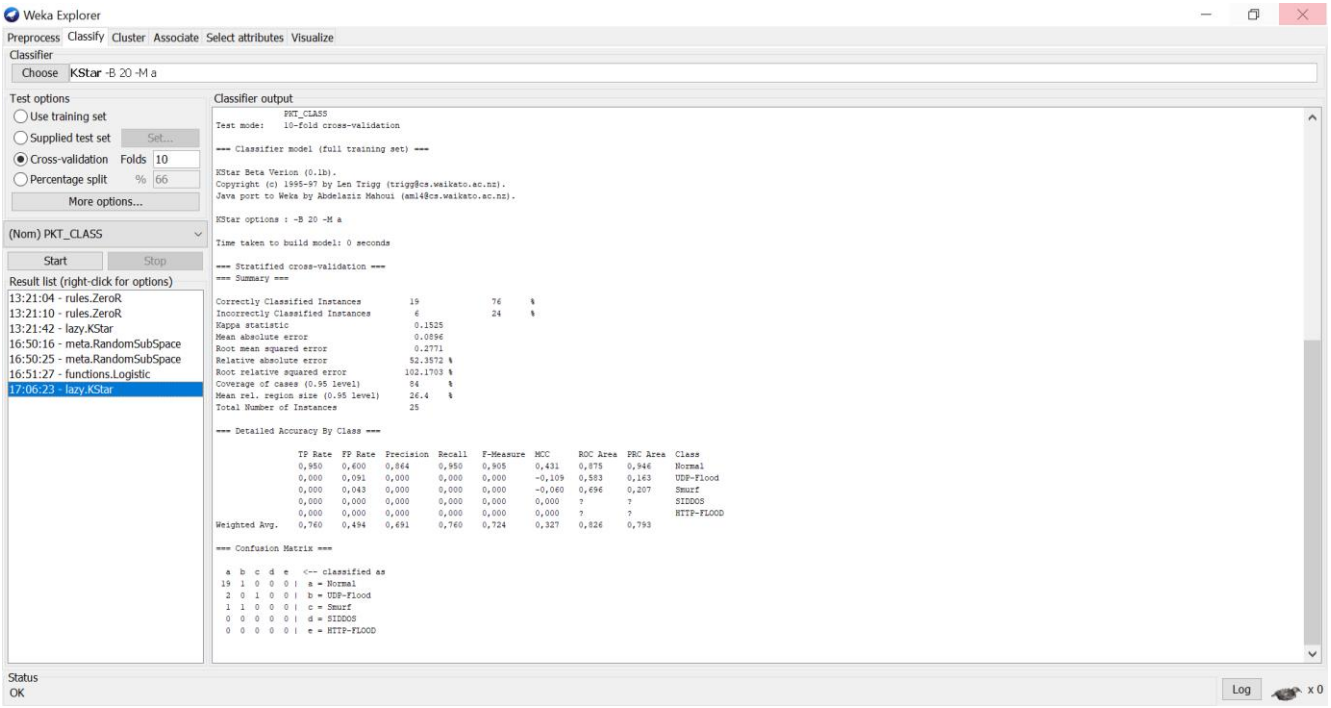


Рисунок 3.4 – Результати класифікації на основі алгоритму логістичної регресії

Отримані результати класифікації класифікації на основі алгоритму логістичної регресії, надають інформацію про те, що правильно класифіковані всі вибірки мережного трафіку (25 випадків). Отримані результати класифікації наведені в табл. 3.2.

Таблиця 3.2 – Результати класифікації на основі алгоритму логістичної регресії

DDoS-атака	Normal	UDP-flood	Smurf	SI DDoS	HTTP-flood
Normal	14	2	4	0	0
UDP-flood	0	0	3	0	0
Smurf	0	1	0	0	0
SI DDoS	0	0	0	0	0
HTTP-flood	0	0	0	0	0

Класифікація на основі алгоритму алгоритму логістичної регресії полягає у визначенні, чи відноситься дані трафіку до шкідливого або ні. Згідно результатів класифікації наведених в табл. 3.2 можна стверджувати, що дана класифікація трафіку DDoS-атак має похибки.

3.6 Вирішення завдання класифікації алгоритмом випадкового лісу Random forest

Випадковий ліс (Random forest, RF) – це алгоритм машинного навчання, який передбачає використання кількох дерев рішень для виконання завдань класифікації. Усі дерева будуються незалежно за таким алгоритмом [16]:

1. Для кожного n = 1,.. „N генерується вибірка X, за якою будується вирішальне дерево.

2. За заданим критерієм вибирається найкращий ознака, за якою проводиться розбиття дерева (До вичерпання вибірки);

3. Дерево будується доти, доки в кожному з листя буде не більше n. об'єктів або поки не буде досягнуто певної глибини дерева.

4. При кожному розбитті спочатку вибирається m випадкових ознак із n вихідних.

5. Пошук оптимального поділу вибірки продиться тільки серед них.

Вирішальний класифікатор алгоритму випадкового лісу [16]:

$$a(x) = \frac{1}{N}\sum_{i=1}^{N} b_i(x),$$

Таким чином, для завдання класифікації вибирається рішення голосуванням з більшості.

Найчастіше алгоритм випадкового лісу працює краще, ніж одне дерево рішень, не вимагає масштавання даних, проте погано обробляє розрізні та високорозмірні дані [16].

Алгоритм випадкового лісу представлений на рис. 3.5.

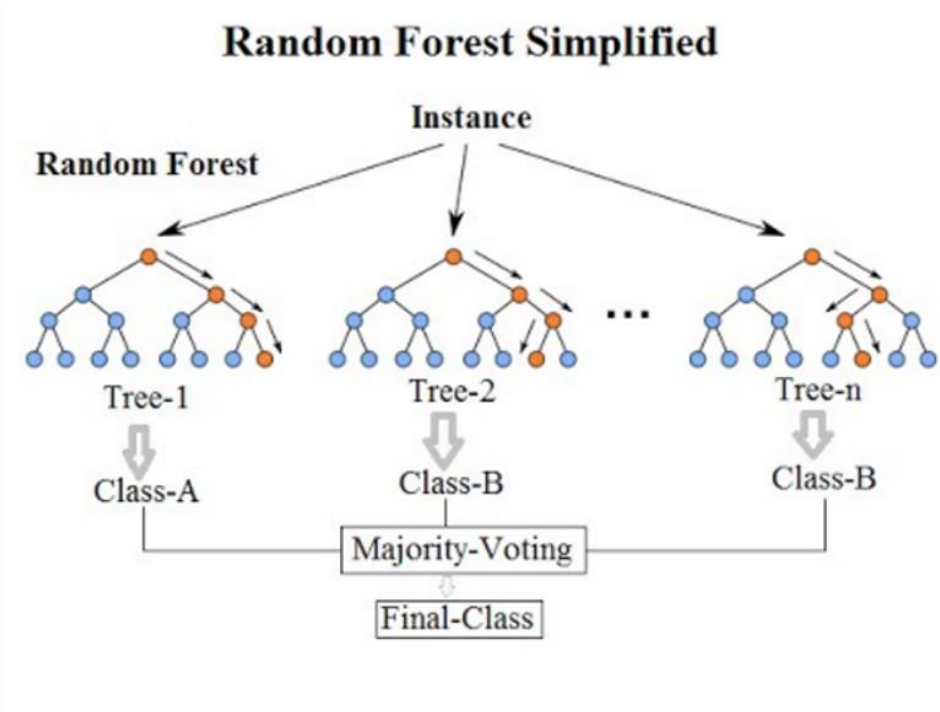


Рисунок 3.5 – Алгоритм випадкового лісу Random forest

Проведена класифікація алгоритмом випадкового лісу Random forest (рис. 3.6) за допомогою пакету Weka 3.7.1 [14], для загальної кількості випадків 25, наведених на рис. 3.1 та 3.2, серед яких, дані трафіку трьох DDoS-атак виду UDP-flood, двох DDoS-атак Smurf та двадцяти атрибутів нормального трафіку без атак.

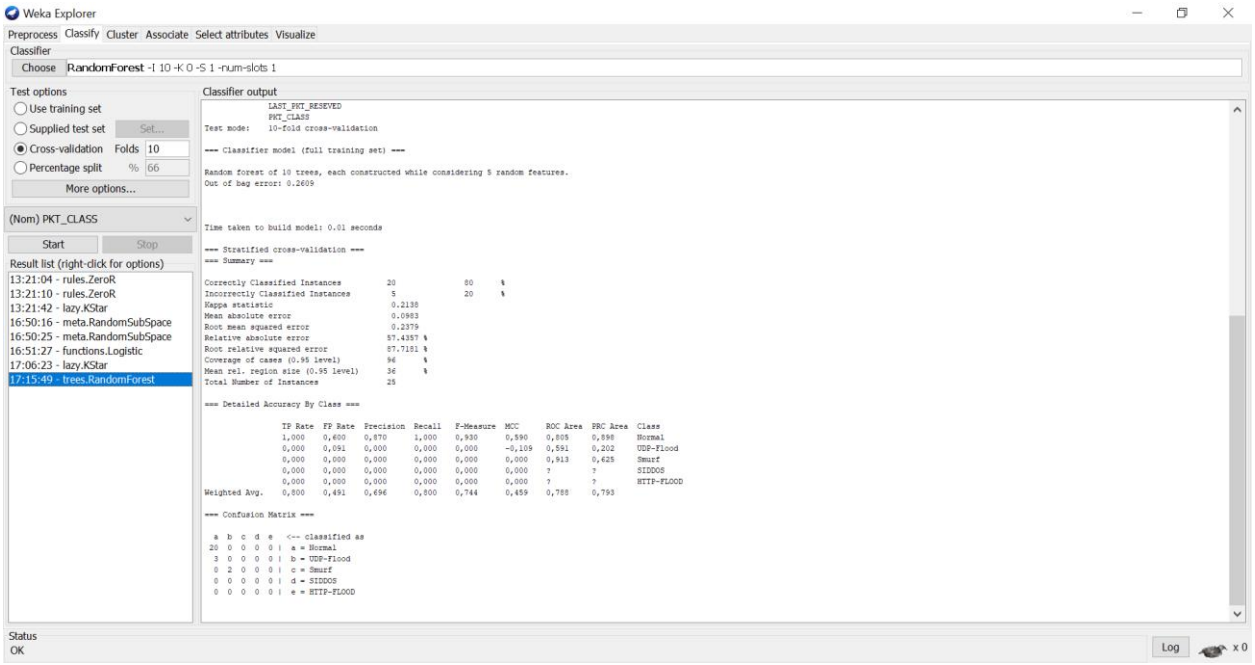


Рисунок 3.6 – Результати класифікації на основі алгоритму випадкового лісу Random forest

Отримані результати класифікації класифікації на основі алгоритму випадкового лісу Random forest, надають інформацію про те, що правильно класифіковані всі вибірки мережного трафіку (25 випадків). Отримані результати класифікації наведені в табл. 3.3.

Таблиця 3.3 – Результати класифікації на основі алгоритму алгоритму випадкового лісу Random forest

DDoS-атака	Normal	UDP-flood	Smurf	SI DDoS	HTTP-flood
Normal	20	0	0	0	0
UDP-flood	3	0	0	0	0
Smurf	0	2	0	0	0
SI DDoS	0	0	0	0	0
HTTP-flood	0	0	0	0	0

Класифікація на основі алгоритму алгоритму на основі алгоритму випадкового лісу Random forest полягає у визначенні виду DDos-атаки. Згідно результатів класифікації наведених в табл. 3.3 можна стверджувати, що дана класифікація трафіку DDoS-атак має похибки.

3.7 Порівняння результатів визначення аномалій тарфіку DDoS-атак

Розглянемо похибку визначення аномалій тарфіку DDoS-атак за допомогою рішення задачі класифікації машинного навчання аномалій трафіку DDoS-атак на основі алгоритму логістичної регресії, алгоритму k-найближчих сусідів та алгоритм випадкового лісу Random forest. Знайдемо значення середньої абсолютної похибки MAPE:

$$\mathrm{MAPE} = \frac{1}{\mathrm{N}} \sum_{\mathrm{t}=1}^{\mathrm{N}} \frac{\left|\mathrm{Z(t)} - \overline{\mathrm{Z}}\mathrm{(t)}\right|}{\mathrm{Z(t)}} \cdot 100\% \qquad (3.4)$$

де Z(t) – фактичне значення показника, Z(t) – визначение за допомогою класифікації та відповідног алгоритму, N – кількість виконаних вимірів.

Таблиця 3.4 – Порівняння точності визначення DDoS-атак за допомогою різних алгоритмів класифікації

Значення середньої абсолютної похибки MAPE,%	Normal	UDP-flood	Smurf	SI DDoS	HTTP-flood
Класифікатор логістичної регресії	59,7%	64,3%	56,7%	-	-
Класифікатор k-найближчих сусідів	42,2%	52,3%	45,7%	-	-
Класифікатора Random forest	11,3%	9,8%	8,7%	-	-

За результатми досліджень можливо зробити наступні висновки, для заданого datasets/ddos-attack-network/ та видів DDoS-атак, які надані у вихідному Datasets (UDP-flood, Smurf, SI DDoS, HTTP-flood та трафіку без атак) найменша похибка класифікації аномалій трафіку DDoS-атак визначається за допомогою класифікатора Random forest – від 8,7% до 11,3% (рис. 3.7) [17].

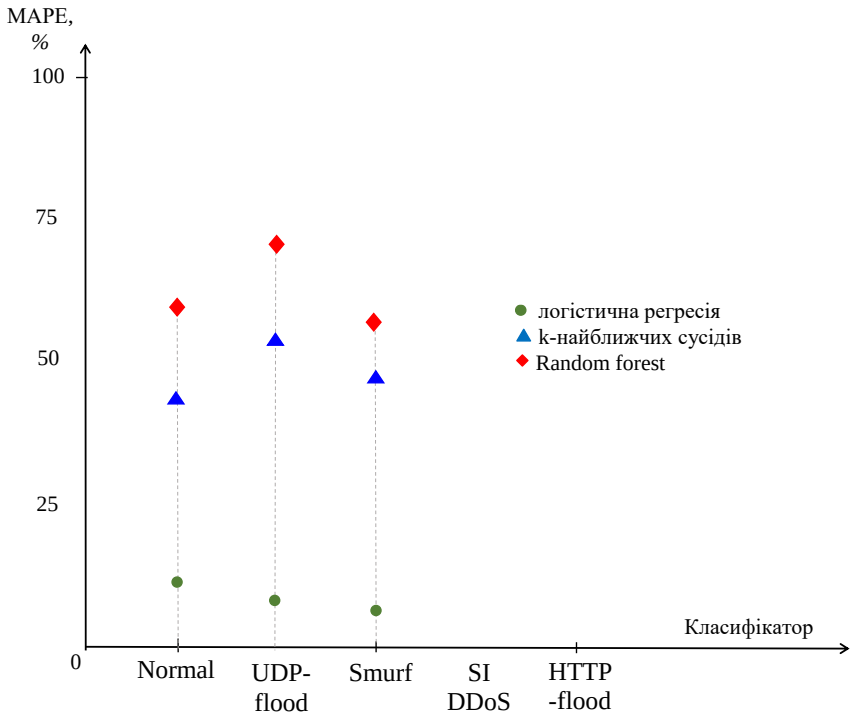


Рисунок 3.7 – Результати точності MAPE визначення DDoS-атак за допомогою різних алгоритмів класифікації: логістичної регресії, методу *k*-найближчих сусідів *ma* випадкового лісу Ramdom forest

Результати дослідження можуть допомогти в розробленні систем класифікації, в яких можна використовувати кілька методів класифікації для підвищення надійності і узгодженості класифікації.

ВИСНОВКИ ТА РЕКОМЕНДАЦІЇ

В роботі проведено дослідження аналіз аномалій мережного трафіку DDoS-атак з метою підвищення точності їхнього детектування.

Результати роботи такі:

1. Проведено класифікацію різних видів DDoS-атак та розглянуті аномалії мережного трафіка DDoS-атак. Проаналізовано основні особливості DDoS-атак та характеристики трафіку атак.

2. Для дослідження використано реальний трафік DDoS-атак з відкритого ресурсу https://www.kaggle.com мережних атак, https://www.kaggle.com/datasets/jacobvs/ddos-attack-network-logs/data, який містить позначені мережеві журнали різних типів мережевих атак, таких як, UDP-Flood, Smurf, SIDDOS, HTTP-FLOOD та звичайний трафік.

3. Запропоновано дослідження аномалій мережного трафіка DDoS-атак на базі статистичних методів та алгоритмів класифікації машинного навчання, таких як, алгоритм логістичної регресії, алгоритм k-найближчих сусідів та алгоритм випадкового лісу.

4. Проведено оцінку точності визначення DDoS-атак за допомогою різних алгоритмів класифікації за допомогою середньої абсолютної похибки MAPE. Встановлено, що найменша похибка класифікації аномалій трафіку DDoS-атак визначається за допомогою класифікатора Random forest – від 8,7% до 11,3%.

5. Отримані результати дозволять забезпечити використання для класифікації аномалій трафіку DDoS-атак алгоритмів, які дозволяють зменшити похибку тим самим збільшивши точність детектування трафіку.

ПЕРЕЛІК ПОСИЛАНЬ

1. ITU-T. Recommendation X.1208. X-series: Data Networks, Open Systems Interconnection and Security, 2019.

2. ITU-T. Recommendation X.1042. X-series: Data Networks, Open Systems Interconnection and Security, Information and Network Security – Network Security, 2019.

3. D.K. Bhattacharyya, J.K. Kalita DDoS Attacks: Evolution, Detection, Prevention, Reaction, and Tolerance, Hapman and Hall/CRC Press, 1st edition, 2016.

4. T. Radivilova, L. Kirichenko, O. Lemeshko and all, "Analysis of anomaly detection and identification methods in 5G traffic," 2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), 2021.

5. Lemeshko O., Yevdokymenko M., Yeremenko M., Kuzminykh I. Cyber Resilience and Fault Tolerance of Artificial Intelligence Systems: EU Standards, Guidelines, and Reports. Proceedings of the Selected Papers on Cybersecurity Providing in Information and Telecommunication Systems (CPITS 2020). Kyiv, Ukraine. CEUR, 2020. P. 99-108.

6. Strelkovskaya I.V. Self-similar traffic in G/M/1 queue defined by the Weibull distribution/ I.V. Strelkovskaya, T.I. Grygoryeva, I.N. Solovskaya // Radioelectronics and Communications Systems. – 2018. – V. 61, № 3 (2018). – P. 173-180.

7. Tabia, K.; Leray, P. Bayesian network-based approaches for severe attack prediction and handling IDSs' reliability. In International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems; Springer: Berlin/Heidelberg, Germany, 2010; pp. 632–642.

8. Sarigiannidis P. Data Traffic Model in Machine to Machine Communications over 5G Network Slicing / P. Sarigiannidis, M. Dighriri, G. Lee, T. Backer, A. S. D. Alfoud// 2016 9th International Conference on Developments in eSystems Engineering (DeSE), Liverpool, UK, 31 Aug.-2 Sept. 2016. – P. 2161-1343. DOI: 10.1109/DeSE.2016.54

9. Farhadi, H.; AmirHaeri, M.; Khansari, M. Alert correlation and prediction using data mining and HMM. *ISeCure* 2011, *3*, 77–101.

10. Pivarníková, M.; Sokol, P.; Bajtoš, T. Early-Stage Detection of Cyber Attacks. Information *2020*, 11, *560*.

11. Alghazzawi, D.; Bamasag, O.; Ullah, H.; Asghar, M.Z. Efficient Detection of DDoS Attacks Using a Hybrid Deep Learning Model with Improved Feature Selection. Appl. Sci. *2021*, *11*, 11634. https://doi.org/10.3390/app112411634

12. Petrik, B.., Dubrovin, V.I. (2021). Detection of DoS attacks in network traffic by wavelet transform. Applied questions of mathematical modelling. 4(1), P. 186-196. https://doi.org/10.32782/KNTU2618-0340/ 2021.4.1.20

13. https://www.kaggle.com

14. https://www.cs.waikato.ac.nz/ml/weka/

15. Стрелковська І.В. Дослідження статистичних характеристик випадкових величин: методичний посібник / І.В. Стрелковська, Л.І. Соколов, О.П. Яковчук. – Одеса: ОНАЗ ім. О.С. Попова, 2004. – 45 с.

16. Giuseppe Bonaccorso Machine Learning Algorithms, Packt Publishing, 2017.

17. Соловська І.М., Севрюков О.В. Дослідження аномалій мережного трафіка DDoS-атак. IX Всеукраїнська науково-практична конференція студентів, аспірантів та молодих учених «Гуманітарний і інноваційний ракурс професійної майстерності: Пошуки молодих вчених»: матеріали конф., 15 грудні 2023 р.: тези доп. – Одеса: МГУ, 2023.

18. Положення факультету кібербезпеки, програмної інженерії та комп’ютерних наук про підготовку та захист кваліфікаційних робіт бакалаврів та магістрів денної та заочної форми навчання: методичний посібник / І.В. Стрелковська, І.М. Соловська, Т.І. Григор’єва, Д.М. Розенвассер. – Одеса: МГУ, 2023. – 45 с.

Додаток А

ПЕРЕЛІК КОПІЙ ДЕМОНСТРАЦІЙНОГО МАТЕРІАЛУ

Магістерська робота на тему:

«Дослідження аномалій мережного трафіка DDoS-атак»

1

Виконав: студ. 2 курсу, групи КТК-2.1маг
Севрюков О.В.

Керівник: к.т.н., доц. Соловська І.М.

Вступ

- Сучасний технологічний розвиток інфокомунікаційних систем та мереж супроводжується постійною необхідністю забезпечення безпеки мережевої інфраструктури від різних атак, в тому числі атак, що призводять до відмови в обслуговуванні DDoS-атак (SYN-Flood, ICMP-Flood, UDP-Flood).

- Класичні методи виявлення DDoS-атак на відмову в обслуговуванні базуються на детектуванні аномалій трафіку, які характерні для кожного виду атак. В той же час, DDoS-атаки рівнів L2-L4 (UDP-Flood, HTTP-Flood, SYN-Flood, ICMP-Flood) переважно генерують значний обсяг трафіку, а DDoS-атаки прикладного рівня L7 не вимагають генерації значного обсягу трафіку, тому атаки такого типу досить складно ідентифікуються звичайними системами. Сучасні заходи для захисту від DDoS-атак часто є неефективними, що викликає значне зростання їхньої кількості. Важливим питанням є дослідження аномалій мережевого трафіку та вибір певного методу для такого дослідження, який би дозволив забезпечити необхідну точність визначення атак.

2

- В цьому змісті й розрахована магістерська робота, яка розглядає можливості дослідження характеристик трафіку DDoS-атак та в подальшому, базуючись на цих результатах, проведення дослідження аномалій мережного трафіка DDoS-атак на базі статистичних методів та методів машинного навчання, таких як: метод логістичної регресії, метод k-найближчих сусідів та метод Random forest.

Мета та методи дослідження

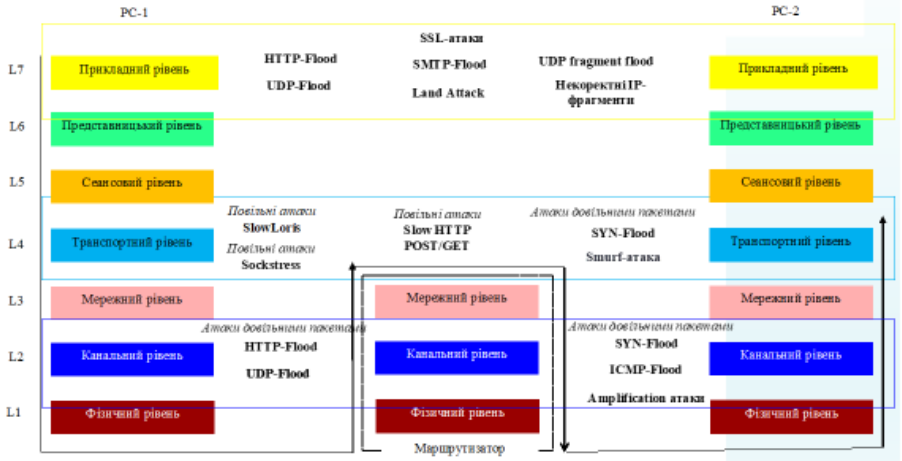
- **Об'єкт дослідження** – мережний трафік DDoS-атак.

- **Мета дослідження** – аналіз аномалій мережного трафіку DDoS-атак з метою підвищення точності їхнього детектування.

3

- Для дослідження запропоновано використання
  – **статистичних методів**
  – **методів машинного навчання на базі класифікаторів з використанням алгоритмів** логістичної регресії, k-найближчих сусідів та випадкового лісу Random forest.

Класифікація DDoS-атак



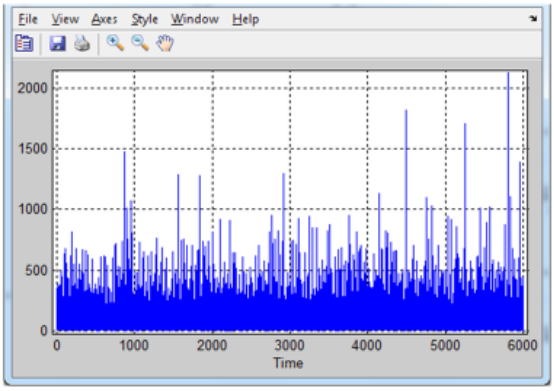
4

Вихідні дані трафіку Dataset/Ddos-attack-network (https://www.kaggle.com)

- мережеві журнали легітимного трафіку та мережевих атак UDP-Flood та Smurf

5

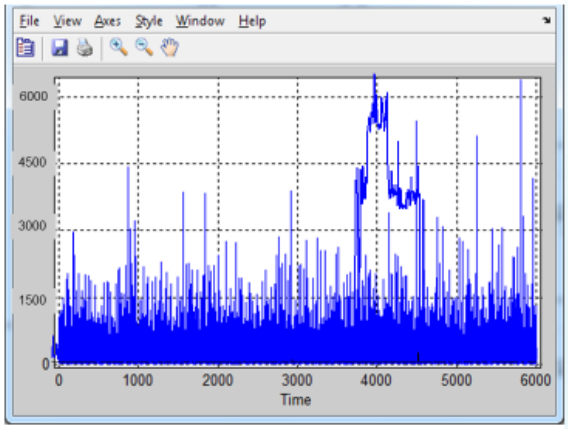
Легітимний трафік мережі (Normal), отриманий за допомогою Dataset/Ddos-attack-network



6

– трафік характеризується значною кількістю «сплесків» інтенсивності пакетів,
– трафік має ознаки самоподібності зі значенням параметра Херста H = 0,8.

Трафік DDoS-атаки виду UDP-flood, отриманий за допомогою Dataset/Ddos-attack-network



7

Статистичний метод виявлення аномалій трафіку DDoS-атаки виду UDP-flood

Таблиця 1 – Розраховані параметри для побудови гістограми трафіку DDoS-атаки

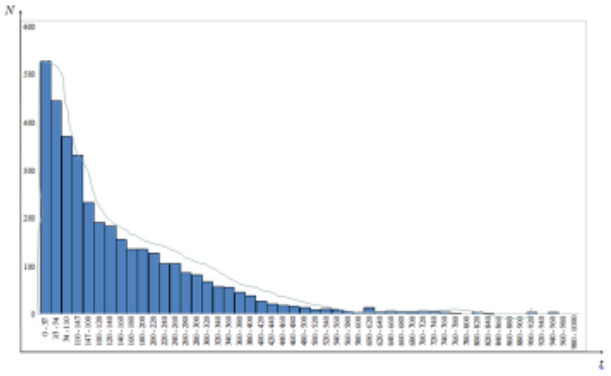
i	1	2	3	4	5	6	7	8	9	10
Інтервали $(x_i; x_{i+1})$	⋮	100-120	120-140	140-160	⋮	700-740	740-760	760-780	⋮	980-1000
Кількість інтервалів m_i	-	6	7	7	-	2	2	1	-	1
Відносна частота p_i	-	0,051	0,050	0,042	-	0,002	0,001	0,003	-	0,000
x_i	-	110,0	130,0	150,0	-	710,0	730,0	770,0	-	990,0
x_i^2	-	12100,0	16900,0	22500,0	-	504200,0	532800,0	592800,0	-	980100,0
$x_i^2 p_i$	-	621,67	847,86	946,11	-	826,62	737,20	165,08	-	0,0

8

Таблиця 2 –Додаткові параметри, необхідні для дослідження статистичних даних трафіку DDoS-атаки

$\widehat{m}_x$	$\widehat{D}_x$	σ	N	μ	Δx	β
159,13	16895,10	119,93	50	0,19	7,7	12,53%

Гистограма трафіку DDoS-атаки



1. Використовуючи дані розподілу вибірки, побудовано емпіричну функцію розподілу пакетів трафіку DDoS-атаки в проміжку [1;1000].
2. По виду цієї кривої зроблено припущення про те, що розподіл виконується за законом розподілу Вейбулла.
3. Використовуючи критерій Пірсона χ^2, перевіряється узгодженість між теоретичним і статистичним розподілами.

9

Використання методів машинного навчання для дослідження аномалій мережного трафіку DDoS-атак (вирішення задачі класифікації)

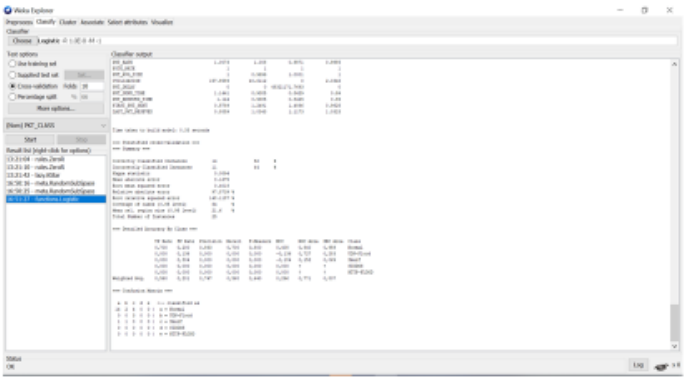
Логістична регресія є моделлю навчання, яка використовується як метод для двійкової класифікації. Логістична регресія моделює ймовірність проблем класифікації з двома можливими результатами і може використовуватися для ідентифікації мережного трафіку як шкідливого (true) чи ні (false).

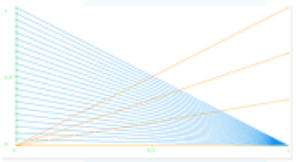
Метод k-найближчих сусідів – дозволяє зробити прогноз для найближчої до існуючої точки навчальної вибірки, знаходячи k найближчих сусідів.

10

Випадковий ліс (Random forest, RF) – це метод, який передбачає використання кількох дерев рішень для виконання завдань класифікації. Алгоритму випадкового лісу є найвикористаним в аналізі кібератак, зокрема ін'єкційних атак; для фільтрації спаму, для виявлення шкідливих програм та іншого.

Результати роботи класифікатора логістичної регресії (Weka 3.7.1)



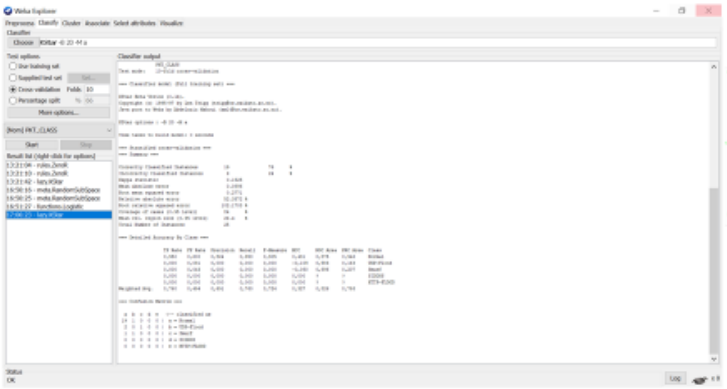


11

Class value UDP-Flood

	Normal	UDP-flood	Smurf	SI DDoS	HTTP-flood
Normal	14	2	4	0	0
UDP-flood	0	0	3	0	0
Smurf	1	1	0	0	0
SI DDoS	0	0	0	0	0
HTTP-flood	0	0	0	0	0

Результати роботи класифікатора k-найближчих сусідів (Weka 3.7.1)

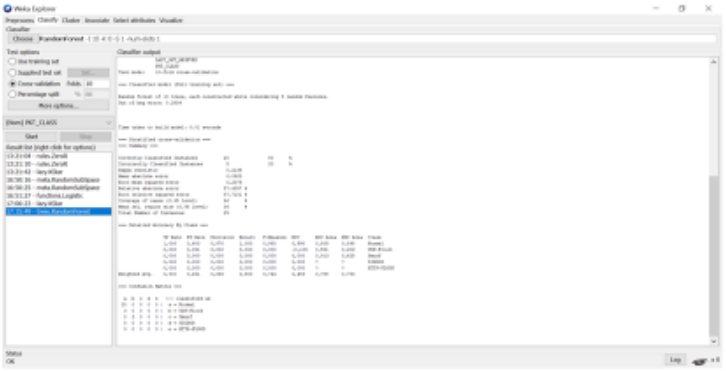


Class value UDP-Flood

12

	Normal	UDP-flood	Smurf	SI DDoS	HTTP-flood
Normal	19	1	0	0	0
UDP-flood	2	0	1	0	0
Smurf	1	1	0	0	0
SI DDoS	0	0	0	0	0
HTTP-flood	0	0	0	0	0

Результати роботи класифікатора Random forest (Weka 3.7.1)



Class value UDP-Flood

13

	Normal	UDP-flood	Smurf	SI DDoS	HTTP-flood
Normal	20	0	0	0	0
UDP-flood	3	0	0	0	0
Smurf	0	2	0	0	0
SI DDoS	0	0	0	0	0
HTTP-flood	0	0	0	0	0

Порівняння точності визначення DDoS-атак за допомогою різних методів класифікації

Значення середньої абсолютної похибки MAPE,%	Normal	UDP-flood	Smurf	SI DDoS	HTTP-flood
Класифікатор логістичної регресії	59,7%	64,3%	56,7%	-	-
Класифікатор k-найближчих сусідів	42,2%	52,3%	45,7%	-	-
Класифікатора Random forest	11,3%	9,8%	8,7%	-	-

14

Висновки

1. Проведено класифікацію різних видів DDoS-атак та розглянуті аномалії мережного трафіка DDoS-атак. Проаналізовано основні особливості DDoS-атак та характеристики трафіку атак.

2. Для дослідження використано реальний трафік DDoS-атак з відкритого ресурсу https://www.kaggle.com мережних атак, https://www.kaggle.com/datasets/jacobvs/ddos-attack-network-logs/data, який містить позначені мережеві журнали різних типів мережевих атак, таких як, UDP-Flood, Smurf, SIDDOS, HTTP-FLOOD та звичайний трафік.

3. Запропоновано дослідження аномалій мережного трафіка DDoS-атак на базі статистичних методів та алгоритмів класифікації машинного навчання, таких як, алгоритм логістичної регресії, алгоритм k-найближчих сусідів та алгоритм випадкового лісу.

4. Проведено оцінку точності визначення DDoS-атак за допомогою різних алгоритмів класифікації за допомогою середньої абсолютної похибки MAPE. Встановлено, що найменша похибка класифікації аномалій трафіку DDoS-атак визначається за допомогою класифікатора Random forest – від 8,7% до 11,3%.

5. Отримані результати дозволять забезпечити використання для класифікації аномалій трафіку DDoS-атак алгоритмів, які дозволяють зменшити похибку тим самим збільшивши точність детектування трафіку.

15